

# Dynamic Trading and Manipulation in Financial Markets

by

Kan Huang

B.A. Mathematics, University of Cambridge (2006)

Submitted to the Sloan School of Management  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

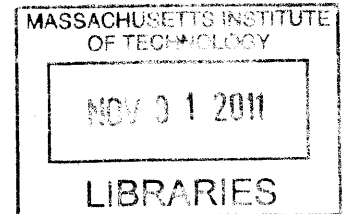
© 2011 Kan Huang. All rights reserved.

The author hereby grants to MIT permission to reproduce and  
to distribute publicly and electronic copies of this thesis document in  
whole or in part.

Author.....  
Sloan School of Management  
26 Apr 2011

Certified by .....  
Jiang Wang  
Mizuho Financial Group Professor  
Thesis Supervisor

Accepted by.....  
Roberto M. Fernandez  
William F. Pounds Professor in Management  
Director, Ph.D. Program, Sloan School of Management



**ARCHIVES**

# Dynamic Trading and Manipulation in Financial Markets

by

Kan Huang

Submitted to the Sloan School of Management  
on 26 Apr 2011, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

Chapter 1 studies how asset managers, due to reputation concerns, manipulate performance through taking latent risk dynamically. It is found that both skilled and unskilled managers load on excessive level of latent risk to boost performance even if investors are fully rational. The equilibrium risk taking by managers has interesting implications on investors' evaluation of manager's skill under normal market conditions and upon crash. Excessive risk taking reduces welfare of investor as well as unskilled managers, which calls for the presence of diligent third-party monitoring. Time required by investors to discover a manager's ability is also significantly lengthened. Our model yields several unique predictions about crash losses, which are supported by empirical analysis using hedge fund data. Besides, it provides complementary explanations for declining returns of large funds and the high demand for structured mortgage securities before the subprime mortgage crisis.

Chapter 2 investigates price manipulation in general equilibrium with the only market imperfection being the presence of a non-competitive large trader. We propose the notion of "pure manipulation", in which the large trader manipulates security prices to improve her welfare but supported by no genuine trading motive. The existence of pure manipulation is equivalent to the failure of the Weak Axiom of Revealed Preference of aggregate security demand at the competitive equilibrium. We state conditions that prohibit pure manipulation. We also demonstrate that heterogeneity in preferences and endowments, large trading needs and remaining insurance demand in the competitive equilibrium could lead to a jointly upward-sloping portfolio demand, which gives rise to pure manipulation that requires arbitrarily small capital commitment. In addition, we establish a link between static and multi-period manipulation and show that dynamic trading reduces manipulation power. Different security structures that complete the markets lead to different equilibrium allocations in the presence of a non-competitive trader.

Chapter 3 analyzes how a risk-averse large institutional investor with price impact trades dynamically in the presence of momentum traders. The larger investor engages in several interesting manipulative behaviors. She may conduct "round-trip" trades to profit from momentum sentiment. She may buy (sell) before planned large sale(purchase) to manipulate intertemporal demand. In addition, she takes profit less aggressively to let the momentum sentiment last longer. Besides, with endogenously generated price impact, we find that higher price volatility does not lead to faster execution.

Thesis Supervisor: Jiang Wang  
Title: Mizuho Financial Group Professor

## Acknowledgmets

I am deeply grateful to my thesis advisor Jiang Wang for his guidance, encouragement and support. I am also very grateful to my committee members Leonid Kogan and Hui Chen for many insightful advices and helpful suggestions. I would like to thank Randy Cohen, Scott Joslin, Andrew Lo, Konstantin Milbradt, Jun Pan as well as other faculty members of the Sloan finance department for invaluable discussions. I would also like to express my gratitude towards my fellow PhD students Jack Bao, Winston Dou, Xing Hu, Brandon Lee, Yichuan Liu, Indrajit Mitra, William Mullins, Zhihua Qiao, Weiyang Qiu, Hong Ru, Yang Sun, Ngoc-Khanh Tran and Amy Zhou for many useful conversations.

I am very much indebted to my family for their love and support.



# Contents

<b>1</b>	<b>Reputation Concerns and Performance Manipulation With Latent Risk</b>	<b>9</b>
1.1	Introduction . . . . .	9
1.2	Related Literature . . . . .	12
1.3	Model Setup . . . . .	13
1.4	Solution of the Equilibrium . . . . .	19
1.4.1	Investors' Learning . . . . .	19
1.4.2	Manager's Maximization . . . . .	21
1.4.3	Equilibrium . . . . .	23
1.5	Analysis of the Equilibrium . . . . .	24
1.5.1	Naive Investors . . . . .	25
1.5.2	Semi-Rational Investors . . . . .	28
1.5.3	Fully Rational Investors . . . . .	29
1.6	Implications of the Equilibrium . . . . .	32
1.6.1	Inefficiencies . . . . .	32
1.6.2	Other Implications . . . . .	39
1.7	Fund Termination . . . . .	40
1.8	Empirical Analysis . . . . .	42
1.8.1	Data . . . . .	45
1.8.2	Idiosyncratic Component . . . . .	45
1.8.3	Testing The Predictions . . . . .	48
1.8.4	Robustness Checks and Discussion . . . . .	49
1.9	Conclusion . . . . .	51
<b>2</b>	<b>Price Manipulation In General Equilibrium</b>	<b>63</b>
2.1	Introduction . . . . .	63
2.1.1	Related Literature . . . . .	65
2.2	Static Economy . . . . .	67
2.2.1	Model Setup . . . . .	67
2.2.2	General Manipulation . . . . .	71
2.2.3	Pure Manipulation . . . . .	74
2.3	Extension To Multi-Period Economy . . . . .	92
2.3.1	Model Setup . . . . .	92
2.3.2	Security Structures . . . . .	94
2.3.3	Pure Manipulation In Commitment Equilibrium . . . . .	98
2.4	Conclusion . . . . .	100

<b>3</b>	<b>Dynamic Trading With Price Impact In The Presence of Momentum Traders</b>	<b>116</b>
3.1	Introduction . . . . .	116
3.2	Model Setup . . . . .	119
3.2.1	Investment Opportunities . . . . .	119
3.2.2	Investors . . . . .	120
3.2.3	Timeline of Trading . . . . .	121
3.3	Equilibrium Definition and Solution . . . . .	122
3.3.1	S-Investors' Problem . . . . .	122
3.3.2	L-Investor's Problem . . . . .	123
3.3.3	Solution of the Equilibrium . . . . .	123
3.4	Equilibrium Results . . . . .	126
3.4.1	All Investors Are R-Investors . . . . .	127
3.4.2	I-Investors Trading Against R-Investors . . . . .	127
3.4.3	R-Investors Trading Against L-Investor . . . . .	128
3.4.4	I-Investors Trading Against L-Investor . . . . .	131
3.5	Conclusion . . . . .	138

# List of Figures

1-1	Optimal Latent Risk Exposure (Against Naive Investors) . . . . .	26
1-2	Effect of Idiosyncratic Volatility (Against Naive Investors) . . . . .	27
1-3	Updating Sensitivity (Against Semi-Rational Investors) . . . . .	28
1-4	Optimal Latent Risk Exposure (Against Semi-Rational Investors) . . . . .	29
1-5	Optimal Latent Risk Exposure (Against Fully Rational Investors) . . . . .	30
1-6	Component in L-type Value Function Due to Reputation . . . . .	31
1-7	Pdf of Post-Crash Reputation when $\hat{p}_t = 0.5$ . . . . .	31
1-8	Effect of Idiosyncratic Volatility (Against Fully Rational Investors) . . . . .	32
1-9	Welfare Loss In Terms of AUM Reduction For Managers . . . . .	34
1-10	Welfare Loss In Terms of AUM Reduction For Passive Investors . . . . .	36
1-11	Minimal Level of Reputation Required . . . . .	39
1-12	Minimal Level of Reputation Required . . . . .	41
2-1	Complete Insurance . . . . .	81
2-2	Numerical Illustration With General Utility Functions . . . . .	85
2-3	CRRA Utilities . . . . .	89
3-1	$E[Q_t]$ and $Var_t[Q_{t+1}]$ for I-Investors v.s. R-Investors . . . . .	128
3-2	Autocorrelation for I-Investors v.s. R-Investors . . . . .	129
3-3	L-Investor's Holding Trajectories After A Position Shock . . . . .	133
3-4	R-Investor's Holding Trajectories After A Position Shock . . . . .	134
3-5	$Corr(Trade_t, \Delta P_t)$ . . . . .	134
3-6	Trajectories After A Dividend Shock, With L-Investor . . . . .	135
3-7	Trajectories After A Dividend Shock, With R-Investors . . . . .	136
3-8	Coefficient on $L_t$ in L-Investor's Strategy . . . . .	136
3-9	Expected Excess Return and Conditional Variance . . . . .	137
3-10	Coefficients On Multiple Past Excess Returns . . . . .	138

# List of Tables

1.1	Expected Discovery Time . . . . .	37
1.2	Dow Jones Credit Suisse Indices Return . . . . .	42
1.3	Hedge Fund Benchmark Factors . . . . .	44
1.5	Seven Factor(plus Lagged Mkt) Regression . . . . .	47
1.6	Large Residual Tests( > 3 Residual Vol) . . . . .	50
1.7	Different Cutoffs For Large Residuals . . . . .	51
1.8	"Live" and "Graveyard" Subsamples . . . . .	52
2.1	Efficiency Implications . . . . .	73
2.2	Dynamic Equilibrium Under Various Complete Markets Security Structures	98

# Chapter 1

## Reputation Concerns and Performance Manipulation With Latent Risk

### 1.1 Introduction

Since the demise of LTCM in 1998, strategies that generate steady returns under normal market conditions but incur large losses upon occasional crashes have caught researchers and practitioners' attention. The "Quant Meltdown"<sup>1</sup> in 2007 provides another vivid example: several popular statistical arbitrage strategies, which were supposed to yield "arbitrage-like" returns with relatively small risk most of the time, suffered tremendous losses when the whole sector was caught in a wave of unwinding. In 2008, people witnessed huge losses from senior tranches of mortgage securities and credit derivatives, which had delivered stable returns before the recession and the collapse of house price. These events beg for the same question: why causes the proliferation of securities or strategies containing significant amount of risk that is latent in nature? While it could be potentially due to neglected disaster risk as argued in Gennaioli, Shleifer and Vishny (2011), the collapse of JWM Partners, a second fund set up by the former LTCM founders, suggests the possibility of alternative explanations. This paper attempts to address the question from the perspective of reputation concerns and associated fund flow.

As Mark Hurley from Goldman Sachs once wittily remarked, "the real business of money management is not managing money, it is getting money to manage." Indeed, an important factor to the success of asset management is to grow asset under management (AUM) and collect fee on a larger base. Several papers<sup>2</sup> have empirically documented that money flows in or out of funds after good or bad performance. As Berk and Green (2004) demonstrated, such flows can be explained by investors' update of managers' ability after observing fund

---

<sup>1</sup>Several famous quantitative players such as Morgan Stanley's Process-Driven Trading group and Goldman Sachs's Global Equities Opportunities Fund posted large losses.

<sup>2</sup>Chevalier and Ellison (1997), Sirri and Tufano (1998) have studied mutual fund data and found that fund flow has been strongly correlated with past returns.

performance, which we shall refer to as "reputation". Reputation concern creates an implicit incentive for managers, which induce them to behave inefficiently. This paper focuses on the effect of reputation concerns on risk taking in the presence of securities or strategies that deliver steady positive returns under normal market conditions but incur potentially large losses upon market or liquidity crashes. We shall refer to this distinct type of risk-return profile as "latent risk", which generalizes over strategies that are often referred to as "insurance selling" or "nickel-picking".

There are several sources that give rise to latent risk. Firstly, it could be simply due to the pay-off profile. For example, credit default swaps and out-of-the-money put, by design, contains significant amount of latent risk. So do structured products such as senior tranches of mortgage securities, which are termed as "economic catastrophe bonds" in Coval, Jurek and Stafford (2008). Secondly, latent risk could arise endogenously for well-known strategies. An inherently good strategy that generates steady return with little risk naturally attracts many fund managers. When a large player of such a strategy is hit, she is forced to liquidate to meet the margin constraint. This might engender the "margin spiral"<sup>3</sup>, which underlies the "Quant Meltdown" and the occasional bad performances of carry trade. Finally, latent risk could be caused by illiquid assets such as bespoke synthetic CDOs that are marked up steadily but register a sharp fall in value when there is a liquidity crash.

Latent risk is of particular interest in the context of performance manipulation because its unique risk-return profile allows managers to boost performance without getting detected until a rare crash. In this paper, we study the optimal exposure of latent risk taken by risk-averse asset managers in a Bayesian rational equilibrium with information asymmetry. An asset manager running a certain type of fund follows a pre-specified strategy stated in the fund prospectus, the risk-return profile of which is known to the investors. A manager could be skilled or unskilled, which is only known to herself but not to investors. Skilled manager can generate an alpha in excess to the pre-specified strategies' expected return. In addition to following the pre-specified strategy, manager can load secretly on latent risk, which boost instantaneous return with little volatility at the expense of a large loss upon crash. Investors cannot observe the level of latent risk-taking by manager but they can form conjecture about it. What they are able to observe is the overall fund returns. Thus, they have to constantly update the reputation<sup>4</sup> of a manager during normal times and upon crash in a Bayesian manner. With fund flow increasing in reputation, it becomes an important concern for managers.

While it is expected that managers without skill choose to manipulate performance by taking excessive (relative to the pure risk-return optimal) latent risk exposure, managers with skill also engage in such behavior, which is puzzling given that they possess genuine alpha-generating ability. The reason is simple: since the unskilled manager will never attempt to generate higher expected return than a skilled one, the equilibrium updating rule by investors always rewards higher realized return with higher reputation. Unable to

---

<sup>3</sup>See Brunnermeier and Pederson (2008).

<sup>4</sup>i.e. the posterior probability of a manager being skilled.

convince investors about their skill due to information asymmetry, the skilled managers optimally choose excessive latent risk exposure to expedite the type discovery. In general, managers with skill take less latent risk exposure than managers without skill under concave utility. However, interestingly, when a skilled manager's reputation is sufficiently low or on the brink of fund termination, the skilled manager may aggressively take even more latent risk to avoid being nailed down to an unskilled one or forced to shut down after a few unfavorable return shocks. Thus, a somewhat unexpected implication is that, when a manager's reputation is already low, a rational investor should not always lower the manager's reputation upon observing a large crash loss.

Rational updating by investors, especially upon crash loss, is very effective at reducing the gap between latent risk exposures taken by skilled and unskilled managers. This is because learning the expected return under Gaussian noise is notoriously slow whereas the information revealed from a crash loss could be substantial. If the unskilled managers take on significantly more exposure than the skilled ones, the gain from boosting the expected return during normal times is outweighed by the loss of reputation revealed by a large loss upon crash. In addition, unskilled managers do not wish to suffer a reputation setback concurrent with large financial loss. These lead to a "noisy pooling" equilibrium, in which the unskilled managers' optimal latent risk exposure is not significantly different from that of the skilled ones. Consequently, the equilibrium evolution of reputation is relatively gradual despite that crash happens rather abruptly. This implies that rational learning alone may not be able to fully explain the drastic withdrawal commonly observed after a fund suffers a substantial loss.

Deviation of latent risk exposure from pure investment optimal creates inefficiency in risk taking. In equilibrium, rational investors can discount the latent risk return from realized performance quite accurately given the "pooling" nature of the equilibrium. As a result, managers cannot gain much reputation although they take excessive latent risk. This calls for the presence of a third-party, who can monitor the risk position of a fund closely and certify to investors that the level of latent risk taking is appropriate. This also justifies the use of investment mandate, which, if effectively enforced, makes performance manipulation through latent risk much more costly. Such self-commitment implementations are in the interest of managers and worth up to 8% of AUM for a manager with mediocre level of reputation. More importantly, the damage to investors' welfare is also substantial. Depending on the level of information asymmetry, the equivalent welfare loss due to excessive latent risk exposure could be up to 7% of money invested. In addition, it is found that latent risk can significantly lengthen the type discovery time by around 30% since higher latent risk exposure taken by unskilled managers in equilibrium reduces the differential expected return and, therefore, the information content of fund performance. This engenders inefficiency in fund flow.

Our model generates a few unique empirical predictions on crash loss, which is related to the level of latent risk exposure. The optimal level depends on the sensitivity of investors' updating rule. Higher updating sensitivity increases the incentive to boost reputation through latent risk exposure. There are a few factors that determine the updating sensitivity. Higher idiosyncratic volatility adds noise to observed fund returns whereas larger

difference in alpha-generating ability between skilled and unskilled managers increases the information content of fund returns. Thus, the former has negative effect and the latter has positive effect on updating sensitivity. Furthermore, when investors are certain of a manager's type (i.e. when reputation is very high or low), they do not weigh much on new observations of fund return, which lowers the sensitivity. Using the TASS dataset for hedge fund returns, we test our predictions and show that the data is consistent with our model predictions.

Besides, previous empirical studies have documented that larger fund has declining return. It is most commonly attributed to trading cost and decreasing return to scale. Our model complements these explanations with a novel one: larger funds tend to be those with high reputation and, as a result, take less latent risk, which reduces the expected return during normal times and overall<sup>5</sup>. Our model also provides an explanation for high demand of mortgage-backed securities, which contributed to the subprime mortgage crisis. Mortgage-backed securities exhibits typical latent risk features. And, with liquidity awash during the period of 2002-2006, new funds were set up and they demanded latent risk. Also, as the overall prior reputation decreased due to limited supply of skilled managers, this further pushed up the demand for latent risk as lower reputation would increase latent risk exposure in general.

This paper is organized as follows: Section 1.2 discusses the related literature. Section 1.3 presents the model setup. Section 1.4 discusses the solution of the equilibrium. Section 1.5 focuses on the analysis of latent risk exposures taken by managers in equilibrium. Section 1.6 considers the implications of the equilibrium. Section 1.7 considers a impact on latent risk exposure if fund termination at low level of reputation is introduced. Section 1.8 tests several implications empirically. Section 1.9 concludes. Proofs and numerical procedures are contained in the Appendix.

## 1.2 Related Literature

With the rise of hedge fund industry, the presence of latent risk, which is easily accessible to hedge fund managers, has attracted people's attention. Malliaris and Yan (2010), Makarov and Plantin (2010) as well as He and Xiong (2010) have approached this issue from different perspectives. Our paper is closely related to Malliaris and Yan (2010), which also studies the excessive deployment of latent risk strategy due to reputation concerns in a discrete-time framework. While they only allow for binary choice of latent risk exposure, we analyze the case where manager can choose the optimal level of exposure from a continuum. This allows us to study the comparative statics and generate testable predictions. Moreover, the equilibrium implication on investors' learning process is different. In their model, investors update positively on a manager's reputation regardless of her true type most of the time until a crash, which is accompanied by a large negative update. In contrast, investors in our model update on the manager's type in the correct

---

<sup>5</sup>Crash risk carries high premium. The overall expected return of latent risk is positive.



direction gradually over time. Makarov and Plantin (2010) finds that the optimal method to game convex compensation contract is through taking latent risk and considers various contracts with commitment to address this issue. He and Xiong (2010) discusses the design of optimal investment mandate to restrict the manager from taking unnecessary latent risk.

Our paper belongs to the growing literature studying the impact of manager's reputation concerns on investment decisions (e.g., Stein and Scharfstein (1990), Dow and Gorton (1997), Dasgupta and Prat (2006, 2008), Vayanos and Woolley (2008)). These studies focus on how reputation concern distorts the use of private information for trading. Our paper focuses on the effect of an outside investment opportunity, namely latent risk, on risk taking.

More broadly, our paper is also related to the large body of literature on managerial incentive that either takes the form of compensation contract (e.g. Jensen and Meckling (1976), Carpenter (2000), Ross (2004), Panageas and Westerfield (2009)) or explicitly specified fund flow function (e.g. Basak, Pavlova and Shapiro (2007, 2008), Basak and Makarov (2009), Chapman, Evans, Xu (2009)) in partial equilibrium. Several papers have also considered the asset-pricing implications in a general equilibrium framework (Arora, Ju and Ou-Yang (2006), Cuoco and Kaniel (2010), Guerrieri and Kondor (2010), Kaniel and Kondor (2010)). Optimal contract has been designed under specific scenarios (e.g. Ou-Yang (2003), Cadenillas, Cvitanic and Zapareto (2007), Dybvig, Farnsworth, Carpenter (2010)) to handle incentive problems.

Finally, performance manipulation against common measures such as Jensen's  $\alpha$  or Sharpe ratio has been analyzed in Goetzmann, Ingersoll, Spiegel and Welch (2007) and Guasoni, Huberman and Wang (2010). Lo (2001) also considers a simple option-writing strategy that delivers high Sharpe ratio. Our paper adds to this branch of literature by studying the effect of latent risk, which has the unusual negatively-skewed risk-return profile, and endogenizing the performance measure through rational Bayesian updating.

### 1.3 Model Setup

In this section, we shall formulate our model in a continuous-time dynamic setup. We shall specify the details of the investment opportunities faced by a fund manager and investors' learning about manager's ability.

#### Pre-specified Strategy

A fund manager has expertise in a particular strategy that falls into some well-known investment style. She states her expertise in the fund prospectus and follows the strategy of her specialty. We shall refer to this strategy as the "pre-specified strategy". Let  $R_t$  denote the the cumulative fund return per dollar (after fee). Equivalently, it is the value at time  $t$  of an \$1 investment in the fund made at time 0. Under the pre-specified strategy,

we assume that  $R_t$  follows a geometric brownian motion with jump

$$\frac{dR_t}{R_t} = \bar{\mu}dt + \sigma dB_t + (e^{-L-\varepsilon} - 1) dN_t \quad (1.1)$$

$\bar{\mu}$  is the expected return (after fee) of the pre-specified strategy. We assume that  $\bar{\mu}$  is known to investors since they are familiar with the investment style, which determines the expected return of the pre-specified strategy.  $\sigma$  is the volatility of the fund return and  $B_t$  is a standard Brownian motion.  $\sigma B_t$  captures the inherent Gaussian risk in the pre-specified strategy. It could be a combination of both systematic risks and idiosyncratic risk. As investors can disentangle the systematic risk component perfectly in continuous-time by comparing the fund return process with the return processes of systematic risk factors, loadings on systematic risk factors play no role in investors' learning process. Thus, for simplicity, we shall assume that  $\sigma B_t$  is purely idiosyncratic, which depends on the specific implementation by the manager and cannot be observed directly by investors.

$N_t$  is a standard Poisson process with intensity  $\lambda$ . If  $N_t$  jumps at  $t$ , it represents a market/style-wide crash that affects the pre-specified strategy. It has been well-documented that the overall equity market carries significant crash risk. In addition, the "quant fund crisis" in 2007<sup>6</sup> demonstrates that certain styles of strategies could have substantial crash risk due to the downward margin spiral of market liquidity and funding liquidity. Upon a crash, the fund's cumulative return  $R_t$  falls to a fraction  $e^{-L-\varepsilon}$  of that before the crash:  $R_t = e^{-L-\varepsilon} R_{t-}$  where  $L > 0$  is the mean size (logged) of the crash loss and  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$  is the idiosyncratic jump loss. A market/style-wide crash has different impact on individual funds. While funds belonging to a certain style suffer a loss of  $L$  on average, the extent, to which an individual fund is affected by the market/style-wide crash, depends on what portfolio that particular fund holds at the time of crash. Thus, there is an idiosyncratic component  $\varepsilon$  in the jump loss. Moreover, as market/style-wide crashes are rare and caused by different underlying problems from time to time, we can treat  $\varepsilon$ 's to be I.I.D over time. Since investors are familiar with the investment style that the pre-specified strategy belongs to,  $L$  is known to investors. However, due to its idiosyncratic nature,  $\varepsilon$  is not observable to investors.

### Manager's Skill

We assume that a manager could be skilled or unskilled in the strategy style she specializes in. A skilled manager (referred to the "H-type") adds an alpha to fund return on top of what is generated by the pre-specified strategy whereas an unskilled manager (referred to as the "L-type") adds nothing. Here, we do not model the source of alpha, which could come from superior information or ability to identify potential arbitrage opportunities. Therefore, under the management of a H-type manager, the cumulative return process has higher drift

$$\frac{dR_t}{R_t} = (\alpha + \bar{\mu}) dt + \sigma dB_t + (e^{-L-\varepsilon} - 1) dN_t$$

---

<sup>6</sup>See Lo and Khandani (2007)

whereas, under L-type, the return is the same as that of the pre-specified strategy.  $\alpha > 0$  is the incremental return contributed by a skilled manager.

In general, we shall write the cumulative return process as

$$\frac{dR_t}{R_t} = \mu^i dt + \sigma dB_t + (e^{-L-\varepsilon} - 1) dN_t, i = H, L \quad (1.2)$$

where

$$\mu^i = \bar{\mu} + \alpha^i \text{ and } \alpha^H = \alpha, \alpha^L = 0$$

We assume that manager knows her type whereas investors cannot observe it. This is motivated by the fact that manager has substantial knowledge of how her strategy is conducted and whether it possesses any alpha through simulation and backtesting. Also, she might have extensive previous trading experience before. Of course, it is possible that sometimes a manager is overconfident and believes that she has alpha-generating ability, which, in reality, she does not possess. Nevertheless, it is still quite reasonable to assume that manager has much better knowledge about her skill level than investors. Thus, there is information asymmetry regarding the skill type of a manager.

### Latent Risk

So far, the return of a fund depends on the risk-return characteristic of the pre-specified strategy and the type of the manager, which are completely exogenous. This is similar to Berk and Green (2004), in which manager does not actively control the return process. We shall now introduce latent risk and the manager can control her fund's exposure to the latent risk. Furthermore, we shall assume that her choice of exposure to the latent risk is over a continuous space rather than being confined to a few discrete levels.

Specifically, a unit of latent risk provides extra return of  $\pi dt$  over the time interval  $dt$  during normal market conditions but incurs a loss of  $\Lambda$  percent when crash occurs (i.e. when  $dN_t = 1$ ). We assume that  $\pi \ll \Lambda$  to emphasize on the skewed risk-return profile. Here, there are two implicit assumptions made. Firstly, we assume that crash due to latent risk is concurrent with the market/style-wide crash (both are driven by  $N_t$ ), which is usually true. For instance, if an equity-oriented fund loads on latent risk by selling out-of-the-money index put options, it is hit by latent risk loss exactly at the same time when the overall equity market crashes. Alternatively, if the latent risk is achieved through holding illiquid assets, then the crash loss caused by latent risk is again simultaneous with the liquidity crash. The same holds true for carry trade, which is known for incurring large loss when the whole investment style suffers from occasional massive unwinding<sup>7</sup>. Merger arbitrage is also found to suffer bad performance when the overall market condition turns sour<sup>8</sup>. The second implicit assumption is that latent risk generates positive return and negative jump losses with no volatility. In reality, some assets with latent risk might be marked to market frequently and, as a result, produce small Gaussian shocks under normal market conditions. But the Gaussian volatility generated by latent risk is usually

---

<sup>7</sup>See Brunneimeier, Nagel and Pedersen (2009).

<sup>8</sup>See Mitchell and Pulvino (2002).

negligible compared to the volatility of the pre-specified strategy. Besides, if latent risk is taken through holding illiquid asset, which is not subject to the marked-to-market procedure, we shall not observe much volatility either. We ignore the Gaussian volatility associated with latent risk to focus on its distinctive feature that is different from the usual Gaussian type of risks.

It is worth noting that we should not consider latent risk taking in the narrow context of hedge fund managers only, who admittedly enjoy greater freedom in their choices of risk taking. A general pension or mutual fund manager could also access latent risk through marginal portfolio adjustment. For instance, an international bond fund manager can load on latent risk by reducing the portfolio weight of Japanese government bond and increasing the portfolio weight of Australian government bond. The marginal effect achieved is exactly a carry trade strategy. Similarly, many pension fund managers chose to hold senior tranches of mortgage securities rather than safer treasury bonds, which yielded a steady spread of 1-2% annually during years of economic boom and suffered almost no default loss. However, during the credit crisis, substantial loss was incurred due to collective default by borrowers as documented in Adelino (2009). In short, being able to load on latent risk should be regarded as a general phenomena among fund managers rather than the special privilege of hedge fund managers alone.

Latent risk is of particular interest to managers because usual risk sources (say aggregate equity market risk) generate greater return at the expense of greater volatility. With frequent observations of fund returns by investors nowadays (most commonly daily or weekly), variations in level of volatility can be detected by investors. Thus, if manager loads on the usual sources of risk, higher return will be discounted by investors, who notice the increase in volatility and understand that extra risk have been taken on. Therefore, they will not perceive the manager with higher reputation. Latent risk delivers return with relatively little volatility. Thus, investors cannot detect it unless a rare crash occurs. As argued in Lo (2007), a simple strategy of selling out-of-the-money index put options can deliver a significant positive alpha in an unsophisticated linear regression framework.

Manager of type  $i$  can choose her exposure of latent risk  $\phi_t^i$  at time  $t$  where  $i = H, L$ . By doing so, the instantaneous return during normal time will be incremented by  $\pi\phi_t^i dt$  and the cumulative return process will be

$$\frac{dR_t}{R_t} = (\mu^i + \pi\phi_t^i) dt + \sigma dB_t$$

However, upon a crash, she will lose  $\Lambda\phi_t^i$  more on top of the crash loss incurred by the pre-specified strategy:

$$R_t = \left(1 - \underbrace{\Lambda\phi_t^i}_{\text{Due to latent risk}}\right) \underbrace{e^{-L-\varepsilon}}_{\text{Due to pre-specified strategy}} R_t -$$

Putting together, the overall cumulative return follows

$$\frac{dR_t}{R_t} = (\mu^i + \pi\phi_t^i) dt + \sigma dB_t + [(1 - \Lambda\phi_t^i) e^{-L-\varepsilon} - 1] dN_t \quad (1.3)$$

Investors cannot observe the precise level of latent risk exposure  $\phi_t^i$  although they may form conjectures about it. While SEC requires quarterly disclosure of portfolio holdings with some lag in 13-F filings for funds with more than 100 million under management, investors can infer limited information from it. First of all, managers trade at much higher frequency than quarterly basis. Thus, the 13-F filings provides at most a snapshot of portfolio holdings, which changes on daily, weekly or monthly basis. This might be subject to the "window-dressing" efforts undertaken by fund managers as well. Secondly, short positions and many derivatives contracts are not required to be reported. Thus, investors cannot gather the full information about portfolio position. Last but not least, the source of latent risk can be obscured by the vast number of investment positions and complex financial contracts in the portfolio, which makes the inference problem even harder.

### Learning By Investors

In summary, we have assumed that investors can only observe the cumulative return  $R_t$  but cannot observe:

1. the type of manager
2.  $B_t$  and  $\varepsilon$ , which are idiosyncratic to a fund
3. the choice of latent risk exposure  $\phi_t^i$  by manager

Thus, they try to learn the type of manager after observing the cumulative return from time 0 to  $t$ ,  $R_{[0,t]}$ , and update the probability of the manager being a skilled H-type

$$\begin{aligned} \hat{p}_t &= \Pr(i = H | R_{[0,t]}) \\ &= E_t^I [1_{\{i=H\}}] \end{aligned} \quad (1.4)$$

where  $E_t^I [\cdot]$  denotes the conditional expectation based on investors' information. Since there are only two types of managers,  $\hat{p}_t$  fully captures the reputation of a manager. Thus, we shall refer to  $\hat{p}_t$ , the posterior probability of a manager being skilled, and "reputation" interchangeably.

### Fund Flow and Reputation Concern

We assume that the fund flow rate is an exogenous increasing function  $f(\hat{p}_t)$  to capture the simple fact that fund flows into hands of managers with greater reputation. The purpose of introducing fund flow into our model is to provide a simple but realistic channel for reputation to reward managers through greater asset under management (AUM). The fund flow is gradual in our model. This can be justified by costly search among some investors as argued in Sirri and Tufano (1998). Also, not all investors are actively monitoring the performance of the fund constantly. Moreover, even if investors are completely sure that the manager is of L-type, they may still invest with the manager for diversification purposes. Though a H-type manager generates higher expected return,

there is still room for the presence of L-type manager following the same pre-specified strategy to reduce idiosyncratic risk. Finally, investors are wary of operational risk as pointed out in Brown, Goetzmann, Liang and Schwartz (2008). They are cautious with high concentration and will not delegate all of their assets to a few H-type managers only. This is illustrated by the fact that, when BlackRock acquires Barcap Global Investors, investors fled to avoid concentration<sup>9</sup>. For simplicity, let  $f(\cdot)$  be linear and increasing in reputation  $\hat{p}_t$ . We further assume that  $f(0.5) = 0$  so that higher  $\hat{p}_t (> 0.5)$  induces fund inflow whereas lower  $\hat{p}_t (< 0.5)$  induces fund outflow.

Let  $W_t$  denote the asset under management (AUM) of the fund. With fund flow taken into account, the AUM evolves as follows

$$\begin{aligned} \frac{dW_t}{W_t} &= \frac{dR_t}{R_t} + f(\hat{p}_t) dt \\ &= (\mu^i + \pi\phi_t^i + f(\hat{p}_t)) dt + \sigma dB_t + [(1 - \Lambda\phi_t^i) e^{-L-\varepsilon} - 1] dN_t \end{aligned} \quad (1.5)$$

Part of the growth in AUM comes from investment return. The rest comes from attracting fund flow with a higher level of  $\hat{p}_t$ . Since a manager wishes to have higher AUM, fund flow, which depends on  $\hat{p}_t$ , introduces reputation concerns for the manager.

### Manager's Objective

We assume that managers receive a management fee as a fraction  $\beta$  of AUM  $W_t$ . Managers try to maximize constant relative risk aversion (CRRA) utility over fee received.

$$\max_{\phi_{[t,\infty]}^i} E_t \left[ \int_t^\infty e^{-\rho s} \frac{(\beta W_s)^{1-\gamma}}{1-\gamma} ds \right]$$

Here we assume that managers consume the fee she receives and abstract away the potential consumption-saving decision. Furthermore, as our focus is the effect of reputation concern on level of latent risk exposure, we assume that the fee is a simple fraction of the AUM. In other words, we do not consider the effect of incentive contracts such as symmetric fulcrum fee seen among mutual funds or those involves high-water mark, which is popular among hedge funds. Also, we assume  $\gamma > 1$  to avoid unrealistic excessive risk taking behavior. For CRRA utility,  $\beta$  does not matter. Manager effectively solves

$$\max_{\phi_{[t,\infty]}^i} E_t \left[ \int_t^\infty e^{-\rho s} \frac{W_s^{1-\gamma}}{1-\gamma} ds \right] \quad (1.6)$$

### Definition of Equilibrium

An equilibrium is defined as  $\{\phi_t^H, \phi_t^L, \hat{\phi}_t^H, \hat{\phi}_t^L\}$  where  $\phi_t^i$  is an adapted process denoting the latent risk exposure taken by manager of type- $i$  and  $\hat{\phi}_t^i$  denotes the conjectured latent risk exposure of type- $i$  by investors (who does not observe  $\phi_t^i$ ) such that

---

<sup>9</sup> As reported in Wall Street Journal on Jul 22nd, 2010.

1. Manager of type- $i$ 's latent risk exposure  $\phi_t^i$  solves  $\max_{\phi_{[0,\infty]}^i} E \left[ \int_0^\infty e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} dt \right]$
2. Investors correctly conjecture manager's latent risk exposure

$$\hat{\phi}_t^i = \phi_t^i \quad \forall t, i = H, L$$

and form their belief about manager's type through Bayesian updating:

$$\hat{p}_t = \Pr(i = H | R_{[0,t]})$$

## 1.4 Solution of the Equilibrium

With infinite horizon and CRRA utility, we look for a stationary equilibrium that is Markov with the state variable being  $\hat{p}_t$ . Hence, the choice of latent risk exposure  $\phi_t^H$  and  $\phi_t^L$  should only depend on the reputation at time  $t$ ,  $\hat{p}_t$ . In other words, they are time-invariant. As a result, we may drop the subscript  $t$  and denote the choice of latent risk exposure as  $\phi^H(\cdot)$  and  $\phi^L(\cdot)$ . We shall verify later that this is indeed an equilibrium.

The equilibrium needs to be solved in 3 steps. First, we need to find out how investors update  $\hat{p}_t$  given their conjectures about manager's choice of latent risk exposure. Second, we need to find out what is manager's optimal latent risk exposure  $\phi_t^i$  taking as given the investors' updating rule. Finally, we need to find out the fixed-point equilibrium, in which manager of a given skill type chooses the optimal latent risk exposure and investors correctly conjecture managers' strategies.

### 1.4.1 Investors' Learning

Investors' learning takes two forms: updating under normal market conditions ( $dN_t = 0$ ) and updating upon a crash ( $dN_t = 1$ ). When they update, they have to form a conjecture about manager's strategy  $\hat{\phi}^i$  for  $i = H, L$ . Let the difference between the the investors' conjecture about latent risk exposures of the two types of manager be  $\hat{\Delta} = \hat{\phi}^H - \hat{\phi}^L$ . The following proposition states the Bayesian updating rule based on investors' conjecture.

**Proposition 1** *Conditioning on no jump in time interval  $[t, t + dt]$ , the reputation of a manager evolves as:*

$$d\hat{p}_t = \beta(\hat{p}) \left( \frac{dR_t}{R_t} - E_t^I \left[ \frac{dR_t}{R_t} \right] \right) \quad (1.7)$$

where

$$\beta(\hat{p}_t) = \frac{\hat{p}_t(1 - \hat{p}_t) \left( E_t^I \left[ \frac{dR_t}{R_t} | i = H \right] - E_t^I \left[ \frac{dR_t}{R_t} | i = L \right] \right)}{\sigma^2 dt} \quad (1.8)$$

**Proof.** See Liptser and Shirayev (1977).  
■

$\frac{dR_t}{R_t} - E_t^I \left[ \frac{dR_t}{R_t} \right]$  measures the unexpected return shock for investors, which could be due to the idiosyncratic Brownian shocks or the difference between the actual drift rate of the cumulative return process and that expected by investors. The former carries no information about the manager's skill type. However, investors cannot disentangle the two sources of shock because of information asymmetry.

$\beta$  measures the sensitivity of updating. Note that  $\beta$  is small when  $\hat{p}$  is close to 0 or 1 and big when  $\hat{p}_t$  is around 0.5. When  $\hat{p}_t$  approaches 0 or 1, it indicates that investors are very certain of manager's type and they no longer update much on the new fund returns. At  $\hat{p} = 0$  or  $\hat{p} = 1$ , they are completely certain and there is no updating at all. They are two absorbing states for the reputation process. When  $\hat{p}$  is around 0.5, investors are least certain about the manager's skill type. As a result, the sensitivity of updating is high, which reflects the fact that additional piece of evidence is useful for determining the type of manager. Moreover, the sensitivity  $\beta$  is high for low  $\sigma$ , which is intuitive. The less noise there is, the more information the shock carries. Finally,  $\beta$  depends on  $E_t^I \left[ \frac{dR_t}{R_t} | i = H \right] - E_t^I \left[ \frac{dR_t}{R_t} | i = L \right]$ , which is the difference in expected returns of the two types of manager. This is a measure of how much information content the unexpected shock possesses. In equilibrium (investors correctly conjecture),  $\hat{p}_t$  follows a martingale based on investors' information filtration. However, as the following proposition suggests,  $\hat{p}_t$  is not a martingale from a manager's point of view since she knows her type and the exact drift of  $\frac{dR_t}{R_t}$ .

**Proposition 2** *For manager of type  $i$ , her reputation follows*

$$d\hat{p}_t = \beta(\hat{p}_t) \left( \alpha^i + \pi \phi_t^i + \sigma dB_t - \left[ \hat{p}_t \left( \alpha + \pi \hat{\phi}^H \right) + (1 - \hat{p}_t) \left( \pi \hat{\phi}^L \right) \right] \right) \quad (1.9)$$

and

$$\beta(\hat{p}_t) = \frac{\hat{p}_t(1 - \hat{p}_t) \left( \alpha + \pi \hat{\Delta}(\hat{p}_t) \right)}{\sigma^2} \quad (1.10)$$

When there is a crash, there is a large amount of information revealed, which requires the second type of updating. Consequently, this causes  $\hat{p}_t$  to jump. Suppose the after-jump cumulative return is a fraction  $\chi$  of the before-jump cumulative return:  $R_t = \chi R_{t-}$ . From investors' perspective,  $\chi = \left( 1 - \Lambda \hat{\phi}^i \right) e^{-L - \varepsilon}$ . Without the idiosyncratic loss  $\varepsilon$ , investors will be able to infer the manager's type from the crash loss by backing out the latent risk exposure  $\hat{\phi}^i$ . However, with the idiosyncratic loss, the overall crash loss is stochastic. As a result, they can only perform Bayesian updating.



**Proposition 3** *Upon a crash, investors update according to the following rule:*

$$\begin{aligned}\hat{p}_t &= P^I(\hat{p}_{t-}, \chi) \\ &= \frac{\hat{p}_{t-}}{\hat{p}_{t-} + (1 - \hat{p}_{t-}) \exp \left( \frac{\frac{1}{2\sigma_\varepsilon^2} \ln \left( \frac{1 - \Lambda \hat{\phi}^H(\hat{p}_{t-})}{1 - \Lambda \hat{\phi}^L(\hat{p}_{t-})} \right)}{\left[ \ln \frac{(1 - \Lambda \hat{\phi}^H(\hat{p}_{t-}))(1 - \Lambda \hat{\phi}^L(\hat{p}_{t-}))}{\chi^2} - 2L \right]} \right)}\end{aligned}\quad (1.11)$$

This has an interesting feature. Rational updating by investors may not imply an increase in post-crash reputation  $\hat{p}_t$  upon seeing a higher fraction of wealth preserved (i.e. smaller loss) after jump if H-type manager is conjectured to have higher latent risk exposure:

$$\frac{\partial P^I}{\partial \chi} < 0 \text{ if } \hat{\phi}^H(\hat{p}_{t-}) > \hat{\phi}^L(\hat{p}_{t-}).$$

This is reasonable because if H-type is expected to load more on latent risk at  $t$ , then her loss upon crash at  $t$  is expected to be higher. Thus, higher loss (i.e. lower  $\chi$ ) indicates a higher probability of being skilled.

By law of iterated expectation, for investors,  $\hat{p}_t$  is still a martingale with stochastic jump:  $E_t^I[P(\hat{p}_{t-}, \chi)] = \hat{p}_{t-}$ . Manager can decompose  $\chi$  into component from her latent risk-taking  $(1 - \Lambda \phi_{t-}^i)$  and component from pre-specified strategy  $e^{-L-\varepsilon}$ . Thus, from the perspective of the manager, the updating rule by investors becomes

$$\begin{aligned}\hat{p}_t &= P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon) \\ &= \frac{\hat{p}_{t-}}{\hat{p}_{t-} + (1 - \hat{p}_{t-}) \exp \left( \frac{\frac{1}{2\sigma_\varepsilon^2} \ln \left( \frac{1 - \Lambda \hat{\phi}^H(\hat{p}_{t-})}{1 - \Lambda \hat{\phi}^L(\hat{p}_{t-})} \right)}{\left[ \ln \frac{(1 - \Lambda \hat{\phi}^H(\hat{p}_{t-}))(1 - \Lambda \hat{\phi}^L(\hat{p}_{t-}))}{(1 - \Lambda \phi_{t-}^i)^2} + 2\varepsilon \right]} \right)}\end{aligned}\quad (1.12)$$

### 1.4.2 Manager's Maximization

Manager of type  $i$  takes investors' conjectures  $\hat{\phi}_t^i$  as given and solves

$$\max_{\phi_{[t, \infty]}^i} E_t \left[ \int_t^\infty e^{-\rho s} \frac{W_s^{1-\gamma}}{1-\gamma} ds \right]$$

subject to

$$\frac{dW_t}{W_t} = (\mu^i + \pi\phi_t^i + f(\hat{p}_t)) dt + \sigma dB_t + [(1 - \Lambda\phi_t^i) e^{-L-\varepsilon} - 1] dN_t \quad (1.13)$$

$$d\hat{p}_t = \beta(\hat{p}_t) \left( \alpha^i + \pi\phi_t^i + \sigma dB_t - \left[ \hat{p}_t (\alpha + \pi\hat{\phi}^H) + (1 - \hat{p}_t) (\pi\hat{\phi}^L) \right] \right) + (P(\hat{p}_{t-}, \phi_t^i, \varepsilon) - \hat{p}_{t-}) dN_t \quad (1.14)$$

where first equation describes how AUM evolves according to the risk-return profile of the pre-specified strategy the manager follows, the choice of latent risk exposure  $\phi_t^i$  as well as the fund flow  $f(\hat{p}_t)$ . The second equation describes the evolution of the manager's reputation under investors' conjecture  $\hat{\phi}^i(\cdot)$ .

**Proposition 4** (1) Manager's optimal latent risk exposure  $\phi_t^i$  is a function of  $\hat{p}_t$  and invariant over time.

(2) The value function of type  $i$  manager is of the form  $V^i(W_t, \hat{p}, t) = \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U^i(\hat{p}_t)$

(3)  $\phi^i(\hat{p}_t)$  and  $U^i(\hat{p}_t)$  satisfy the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} 0 = & \min_{\phi^i} \rho + \left[ -\rho - \lambda + (1 - \gamma) \left( \mu^i + \pi\phi^i + f(\hat{p}_t) - \frac{1}{2}\gamma\sigma^2 \right) \right] U^i \\ & + \beta \left\{ (\alpha^i - \alpha\hat{p}_t) + (1 - \gamma)\sigma^2 + \pi \left[ \phi^i - \left( \hat{p}_t \hat{\phi}^H(\hat{p}_t) + (1 - \hat{p}_t) \hat{\phi}^L(\hat{p}_t) \right) \right] \right\} U_{\hat{p}}^i \\ & + \frac{1}{2}\sigma^2\beta^2 U_{\hat{p}\hat{p}}^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda\phi^i)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^i(P(\hat{p}_t, \phi^i, \varepsilon))] \end{aligned} \quad (1.15)$$

with  $\phi_t^i$  minimizing

$$[(1 - \gamma) U^i + \beta U_{\hat{p}}^i] \pi\phi_t^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda\phi_t^i)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^i(P(\hat{p}_t, \phi_t^i, \varepsilon))] \quad (1.16)$$

Given the form of value function and our restriction that  $\gamma > 1$ , we can easily deduce that  $U^i > 0$  since  $V^i$  should be increasing in AUM  $W_t$ , which implies

$$\frac{\partial V^i}{\partial W_t} = \frac{1}{\rho} e^{-\rho t} W_t^{-\gamma} U^i > 0$$

Also, we can conclude that  $U_{\hat{p}}^i < 0$  since  $V^i$  should be increasing in the reputation of the manager

$$\frac{\partial V^i}{\partial \hat{p}_t} = \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U_{\hat{p}}^i > 0$$

The terms to be minimized by  $\phi_t^i$  have very intuitive meaning. Loading on latent risk at level  $\phi_t^i$  could boost the fund return by  $\pi\phi_t^i$ .  $(1 - \gamma) U^i \pi\phi_t^i$  reflects the direct pecuniary benefit from the steady return the latent risk factor provides. Also, with return  $\pi\phi_t^i dt$  higher, her reputation increases by  $\beta\pi\phi_t^i dt$  over the time interval  $[t, t+dt)$ , benefit of which

is reflected in the term  $\beta\pi\phi_t^i U_p^i$ . Since  $(1 - \gamma) U^i < 0$  and  $U_p^i < 0$ , higher  $\phi_t^i$  leads to lower value of the first two terms of (1.16).  $\lambda e^{-(1-\gamma)L} (1 - \Lambda\phi^i)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^i (P(\hat{p}_t, \phi^i, \varepsilon))]$  is a positive term, which counterbalances the linear negative effect of the first two terms and represents the cost of latent risk.  $\lambda (1 - \Lambda\phi^i)^{1-\gamma}$  measures the financial loss (after adjusted for risk-aversion) due to latent risk exposure when crash arrives (with intensity  $\lambda$ ). This is combined with the loss due to pre-specified strategy  $e^{-(1-\gamma)L}$  (adjusted for risk-aversion) and expected investors' updating  $E_t [e^{-(1-\gamma)\varepsilon} U^i (P(\hat{p}_t, \phi_t^i, \varepsilon))]$  upon a crash.

### 1.4.3 Equilibrium

If  $\hat{p}_t = 0$  or 1, when  $\hat{p}_t$  will no longer change since investors are already completely certain of the types. So there is no reputation concern at all. Managers will simply choose the optimal latent risk exposure based on the pure risk-return trade-off. In other words, they will grow the AUM in an "organic" manner.

**Proposition 5** *If  $\hat{p}_t = 0$  or 1,*

$$\phi_t^i = \bar{\phi} = \frac{1 - e^{\frac{\gamma-1}{\gamma}L + \frac{(\gamma-1)^2}{2\gamma}\sigma_\varepsilon^2} \left(\frac{\lambda\Lambda}{\pi}\right)^{\frac{1}{\gamma}}}{\Lambda} \quad (1.17)$$

$$U^i(0) = \frac{\rho}{\rho + \lambda + (\gamma - 1) (\mu^i + \pi\bar{\phi} + f(0) - \frac{1}{2}\gamma\sigma^2) - \frac{\pi}{\Lambda} (1 - \Lambda\bar{\phi})} \quad (1.18)$$

$$U^i(1) = \frac{\rho}{\rho + \lambda + (\gamma - 1) (\mu^i + \pi\bar{\phi} + f(1) - \frac{1}{2}\gamma\sigma^2) - \frac{\pi}{\Lambda} (1 - \Lambda\bar{\phi})} \quad (1.19)$$

*with a parametric restriction that*

$$\rho + \lambda + (\gamma - 1) \left( \bar{\mu} + \pi\bar{\phi} + f(0) - \frac{1}{2}\gamma\sigma^2 \right) - \frac{\pi}{\Lambda} (1 - \Lambda\bar{\phi}) > 0$$

Note that without reputation concern, the level of latent risk-taking is constant and the same for both types. It is straightforward to see that the exposure increases in  $\frac{\pi}{\lambda\Lambda}$ , which is the premium of latent risk. Also, under the parametric assumption that  $\gamma > 1$ , it is decreasing in the mean logged loss of the typical strategy  $L$ . This is understandable since manager is quite risk-averse ( $\gamma > 1$ ). Given a loss will be incurred by the pre-specified strategy when a crash arrives, it is undesirable to aggravate the situation by incurring further loss from latent risk exposure. Finally,  $\bar{\phi}$  is decreasing in  $\sigma_\varepsilon$ , the idiosyncratic jump loss due to risk aversion.

**Proposition 6** *In equilibrium,  $\hat{\phi}_t^i(\hat{p}_t) = \phi_t^i(\hat{p}_t)$ . Define  $\Delta = \phi_t^H(\hat{p}_t) - \phi_t^L(\hat{p}_t)$ . So the above equations simplifies to a fixed-point problem of differential-integral equations*

$$\begin{aligned} 0 = & \rho + \left[ -\rho - \lambda + (1 - \gamma) \left( \mu^H + \pi \phi^H + f(\hat{p}_t) - \frac{1}{2} \gamma \sigma^2 \right) \right] U^H \\ & + \beta \{ \alpha (1 - \hat{p}_t) + (1 - \gamma) \sigma^2 + \pi (1 - \hat{p}_t) \Delta(\hat{p}_t) \} U_{\hat{p}}^H + \frac{1}{2} \sigma^2 \beta^2 U_{\hat{p}\hat{p}}^H \\ & + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^H)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^H (P(\hat{p}_t, \phi^H, \varepsilon))] \text{ for } i = H \end{aligned} \quad (1.20)$$

$$\begin{aligned} 0 = & \rho + \left[ -\rho - \lambda + (1 - \gamma) \left( \mu^L + \pi \phi^L + f(\hat{p}_t) - \frac{1}{2} \gamma \sigma^2 \right) \right] U^L \\ & + \beta \{ -\alpha \hat{p}_t + (1 - \gamma) \sigma^2 - \pi \hat{p}_t \Delta(\hat{p}_t) \} U_{\hat{p}}^L + \frac{1}{2} \sigma^2 \beta^2 U_{\hat{p}\hat{p}}^L \\ & + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^L)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^L (P(\hat{p}_t, \phi^L, \varepsilon))] \text{ for } i = L \end{aligned} \quad (1.21)$$

$$\phi^i = \arg \min \pi [(1 - \gamma) U^i + \beta U_{\hat{p}}^i] \phi_t^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^i)^{1-\gamma} E_t [e^{-(1-\gamma)\varepsilon} U^i (P)] \quad (1.22)$$

with boundary conditions at 0 and 1 specified by (1.18) and (1.19).

The fixed-point involves three parties: potential  $H$ -type, potential  $L$ -type as well as the investors. We require both types of managers to maximize their utility while investors correctly conjecture. Due to the non-linearity and the involvement of integral (i.e. the expectations that appear in the equations above), we have to resort to numerical techniques. The details are included in Appendix.

## 1.5 Analysis of the Equilibrium

In this section, we shall analyze the equilibrium solved in the previous section. As the the equilibrium is a functional fixed-point problem involving three parties (potential  $H$ -type, potential  $L$ -type and investors) and two different forms of updating by investors, we shall analyze step by step to see the effect of each component. We start by analyzing an open loop case first – manager facing naive investors, who "naively" believe that managers will simply follow the pre-specified strategy and refrain from loading on excessive latent risk. This case will yield several basic intuitions, which still holds in fully rational equilibrium. Next, we shall consider the case, in which investors are semi-rational in the sense that they conjecture the latent risk exposure  $\phi_t^i$  correctly to perform continuous-time updating but ignore the information content in jump loss. Finally, we consider the case in which fully rational investors conjectures  $\phi_t^i$  correctly and perform updating both during normal market conditions and upon crash. This opens a potentially important channel for investors to learn about manager's skill type. We shall see that the equilibrium result is "noisy pooling", in which the difference in expected crash loss between  $H$ -type and  $L$ -type is small relative to the idiosyncratic noise.

### 1.5.1 Naive Investors

As long as investors do not perform jump updating, which applies for both the naive and semi-rational investors cases, the reputation levels before and after a crash is the same:  $P(\hat{p}_t, \phi_t^i, \varepsilon) = \hat{p}_t$ . From (1.16), we find that the optimal latent risk exposure by manager of type  $i$  is

$$\phi_t^i = \frac{1 - e^{\frac{\gamma-1}{\gamma}L + \frac{(\gamma-1)^2}{2\gamma}\sigma_\varepsilon^2} \left(\frac{\lambda\Lambda}{\pi}\right)^{\frac{1}{\gamma}} \left(1 + \frac{\beta U_p^i}{(1-\gamma)U^i}\right)^{-\frac{1}{\gamma}}}{\Lambda} \quad (1.23)$$

Since  $(1-\gamma) < 1$  and  $U_p^i < 0$ , we have  $\frac{\beta U_p^i}{(1-\gamma)U^i} > 0$ . Comparing this with (1.17), we see that

$$\phi_t^i(\hat{p}_t) \geq \bar{\phi}$$

with equality holds only when  $\hat{p}_t = 0, 1$ . This suggests that managers always take on excessive latent risk exposure than what is optimal without reputation concern. The magnitude of deviation from  $\bar{\phi}$  (excessive latent risk exposure) depends on  $\frac{\beta U_p^i}{(1-\gamma)U^i}$ , which is determined by two factors:  $\beta(\hat{p}_t)$  and  $-U_p^i/U^i$ . As discussed in section 3.1, the former is the sensitivity of investors' updating to unexpected return shocks. From the manager's point of view,  $\beta$  captures the effectiveness of boosting return through putting on more latent risk. The more sensitive investors' updating rule is to fund performance (i.e. higher value of  $\beta$ ), the larger reputational gain can be achieved by incremental return obtained through latent risk exposure. As a result, manager will be more inclined to load on latent risk given the cost of doing so, loss of  $\Lambda$  percent upon a crash, remains the same (since there is no jump-updating).

$\beta$  has an important effect on the variation of level of latent risk exposure with respect to level of reputation  $\hat{p}_t$ . For naive investors,

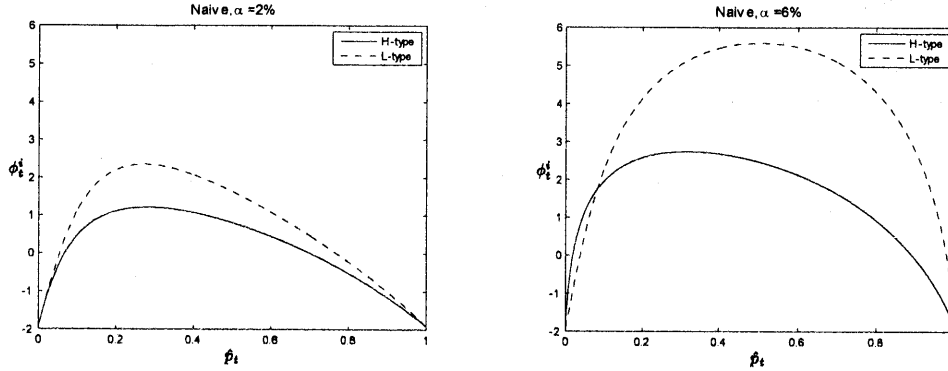
$$\beta(\hat{p}_t) = \frac{\alpha}{\sigma^2} \hat{p}_t (1 - \hat{p}_t)$$

The sensitivity is an inverse-U shaped parabola equal to 0 at the  $\hat{p}_t = 0, 1$  and peaks at  $\hat{p}_t = \frac{1}{2}$ . Intuitively, this makes sense because, at  $\hat{p}_t = \frac{1}{2}$ , investors are most uncertain of the type of manager. They put greatest weight on newly observed unexpected shock of the realized return. At the two ends 0 and 1, investors are completely sure of the manager's type and they stop updating. In general, the closer  $\hat{p}_t$  is to 0 or 1, the more certain investors are about manager's type. Their updating is less sensitive to unexpected shocks of fund return. As a result, the manager has less incentive to take on excessive latent risk when her reputation  $\hat{p}_t$  is very high or low and more incentive when their reputation is mediocre. This suggests an inverse-U shape relationship between the reputation level and the latent risk exposure. The magnitude of  $\beta$  is increasing in  $\alpha$  and decreasing in  $\sigma^2$ . The former measures the difference in expected returns generated by skilled and unskilled managers as perceived by naive investors. The latter is pure idiosyncratic noise, which is unrelated to skill level. Thus, higher  $\alpha$  increases the information content whereas higher  $\sigma^2$  increases the noise level in the unexpected shocks of fund return. As a result, the level of latent risk exposure is high when the skill difference is large and idiosyncratic risk of

the fund return is small.

We see that both types of managers take excessive level of latent risk. While that is expected for the unskilled manager, it is somewhat surprising for the skilled one, who already possess superior alpha-generating skill. This is due to information asymmetry between a skilled manager, who knows her true type, and investors, who cannot tell her types. Unable to convince the investors of skill level, a skilled manager is still incentivized to load on latent risk to speed up the discovery of her superior skill by investors since investors' updating rule always favors higher realized return and imposes no reputational punishment for crash loss (when investors do not perform jump-updating).

Figure 1-1: Optimal Latent Risk Exposure (Against Naive Investors)

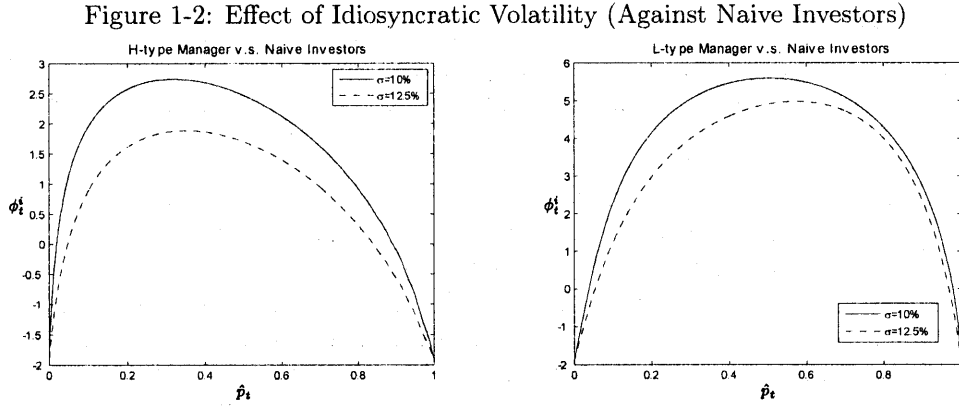


Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

$-U_{\hat{p}}^i/U^i \approx \frac{-\Delta U^i/U^i}{\Delta \hat{p}}$  measures how marginal increase in reputation will lead to a percentage increase in utility. This captures eagerness of a manager to boost her reputation, which contributes to her utility through fund flow. Given that manager's utility is concave, higher fund flow has declining marginal utility for her. Thus,  $-U^i$  is increasing in reputation  $\hat{p}$  at a declining rate. This tends to reduce the manager's eagerness for boosting reputation as her reputation becomes higher. As a result, although the sensitivity  $\beta$  is exactly symmetric about and peaks at  $\frac{1}{2}$ , the declining marginal utility with respect to reputation tends to let the maximum level of latent risk exposure peak at reputation levels less than  $\frac{1}{2}$ . For H-type manager with true alpha, she enjoys higher persistent income through alpha directly, which is equivalent to having a higher level of effective reputation. This reduces her incentive to load on latent risk. Given that both types face the same updating rule by investors, this will in general lead to lower exposure by H-type manager. But this might not be true for very low level of reputation. When her reputation level is close to 0, the proportional marginal utility of reputation might be higher than that of a L-type because, once she is nailed down as an unskilled manager, she forfeits the large

potential benefit of higher fund flow associated with her alpha-generating ability. The potential benefit is especially large if her alpha-generating ability is high, which delivers fast ascent in reputation. As a result, she chooses a higher level of latent risk exposure than a L-type manager in this situation.

Fig 1.1 shows the optimal levels of latent risk exposure taken by skilled and unskilled managers when the alpha generated by skilled type  $\alpha = 2\%$  and  $6\%$ . The optimal from pure risk-return perspective is the level of latent risk exposure at  $\hat{p} = 0, 1$  when there is no reputation concern. As analyzed above, both types of managers take on excessive latent risk exposure. Moreover, we see that the level of latent risk exposure exhibits an inverse U-Shape with respect to reputation regardless of the value of  $\alpha$  and the type of manager. In addition, the maximum level of latent risk exposure is skewed towards the left of  $\frac{1}{2}$ , where updating sensitivity is highest, due to decreasing marginal utility of reputation. This manifests more strongly for the skilled H-type manager. Finally, we see that for most of domain of  $\hat{p}_t$ , L-type takes on more latent risk H-type. In fact, this is true for all levels of reputation when  $\alpha = 2\%$  and investors' updating is not very sensitive. However, when  $\alpha = 6\%$ , the information content is much higher and the sensitivity of investors' updating rule is higher. As a result, H-type takes on more latent risk than L-type does when  $\hat{p}_t$  is sufficiently low to avoid being nailed down as a unskilled manager.



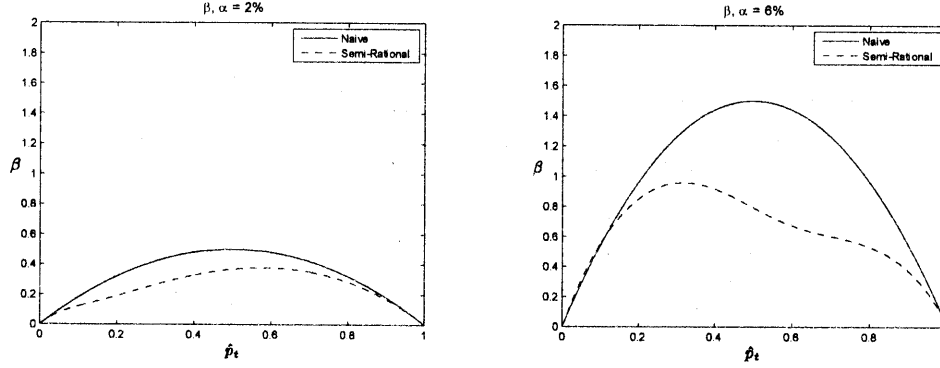
Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

Fig 1.2 demonstrates the effect of idiosyncratic risk  $\sigma$  of the pre-specified strategy on the optimal latent risk exposure taken by H-type and L-type managers. Higher idiosyncratic risk  $\sigma$  increases the noise level in the fund return. As a result, the sensitivity of updating  $\beta$  decreases at all levels of reputation  $\hat{p}_t$ . This creates less incentive for managers to boost return through taking latent risk, which leads to lower levels of optimal latent risk exposure. As  $\sigma$  increases from 10% to 12.5%, we see that both types of managers lower their latent risk exposures.

## 1.5.2 Semi-Rational Investors

In this case, investors are semi-rational. They are fully aware of fact that manager loads on latent risk to boost return and they correctly conjecture the level  $\phi_t^i$ . We assume that they take these into account when performing continuous updating based on observed cumulative return. However, we keep the second channel of updating, which happens upon a crash, shut for the moment. They are semi-rational in the sense that they do perform rational updating under normal market conditions but fail to update  $\hat{p}_t$  upon observing the jump loss during crash. This is a step towards fully rational updating. The situation is no longer open-looped. Rather, manager takes investors' conjecture and their updating rule as given and fully optimizes over the level of latent risk exposure whereas investors conjecture correctly about each type of manager's optimal strategy and form rational updating rule.

Figure 1-3: Updating Sensitivity (Against Semi-Rational Investors)



Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

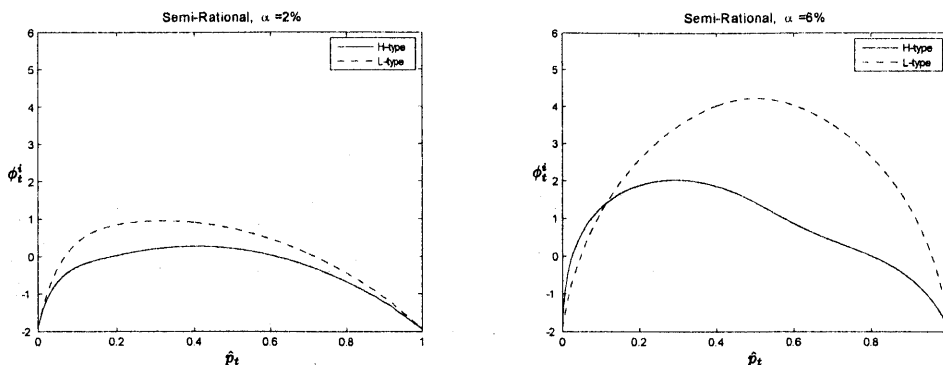
The updating rule by investors is fundamentally changed: once investors take into account of the latent risk-taking  $\phi_t^i$ , the sensitivity  $\beta$  is endogeneously determined

$$\beta(\hat{p}_t) = \frac{\hat{p}_t(1 - \hat{p}_t)(\alpha + \pi\hat{\Delta}(\hat{p}_t))}{\sigma^2} \quad (1.24)$$

where  $\hat{\Delta} = \hat{\phi}_t^H(\hat{p}_t) - \hat{\phi}_t^L(\hat{p}_t)$ . The extra term  $\pi\hat{\Delta}$  reflects the adjustment for difference in conjectured latent risk exposures under H-type and L-type managers. The overall difference in expected fund return is the sum of the difference in alpha, which is the inherent ability of managers, and the difference in levels of latent risk exposure, which is endogenous in the model. Since, for most of the domain of  $\hat{p}_t$ , H-type manager takes less latent risk than L-type manager,  $\hat{\Delta}(\hat{p}_t) < 0$  and  $\beta$  is lowered than the naive case. This is illustrated by Fig 1.3. Lower sensitivity reduces the effectiveness of boosting reputation



Figure 1-4: Optimal Latent Risk Exposure (Against Semi-Rational Investors)



Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

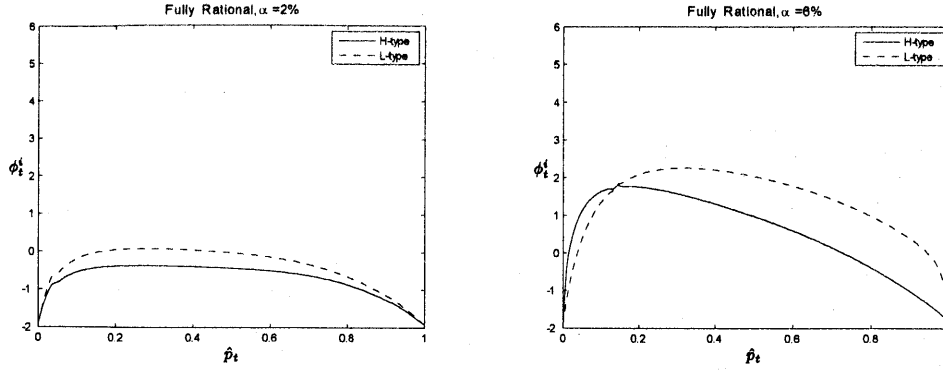
through latent risk and leads to lower latent risk exposure in equilibrium by both types in comparison with the naive case. Fig 1.4 plots latent risk exposure  $\phi^i$  against reputation level  $\hat{p}_t$  for both  $\alpha = 2\%$  and  $\alpha = 6\%$ . We see a decrease in latent risk taking by both type of managers relative to the case, in which investors are naive.

### 1.5.3 Fully Rational Investors

Finally, let us consider fully rational investors who conjecture the latent risk exposures taken by managers correctly and perform rational Bayesian learning both under normal market conditions and upon a crash. As discussed earlier on, crash loss could potentially contain a significant amount of information about the level of latent risk exposure if H-type and L-type's latent risk exposure are significantly different. As shown by Fig 1.1 and Fig 1.4, when  $\alpha$  is high, the difference in latent risk taking by H-type and L-type of managers are large if investors ignore information contained in crash loss (they are naive or semi-rational). However, as we turn on the new channel of learning through crash loss, the equilibrium results are changed dramatically. With a moderate level of idiosyncratic crash loss ( $\sigma_\varepsilon = 20\%$ ), we see that this new channel of learning brings down the level of L-type's latent risk exposure significantly, which is illustrated by Fig 1.5.

The new channel of learning imposes a new reputational cost upon crash for L-type manager if she loads on more latent risk exposure than H-type manager in equilibrium, which leads investors to adjust reputation downward upon observing a large crash loss. It is well-known that learning the drift of a process with Brownian noise is a slow process. In contrast, learning upon crash could be fast if L-type takes considerably more latent risk than H-type, which will be reflected by a much larger crash loss on average. This suggests that the difference between latent risk exposures taken by the two types of managers will be significantly smaller than the difference in the semi-rational case because

Figure 1-5: Optimal Latent Risk Exposure (Against Fully Rational Investors)



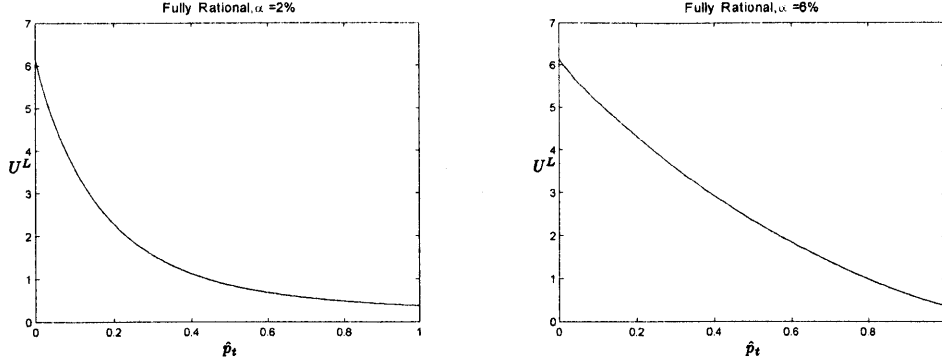
Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

of the reputation cost for the L-type manager due to crash updating. In addition, the convexity of the value function (see Fig. 1.6) as a result of risk aversion plays an important role. Recall that when manager solves for optimal  $\phi_t^i$  in (1.16), the cost of higher latent risk exposure is captured by the term  $\lambda e^{-(1-\gamma)L + \frac{1}{2}(1-\gamma)^2\sigma_\varepsilon^2} (1 - \Lambda\phi^i)^{1-\gamma} E_t[e^{-(1-\gamma)\varepsilon} U^i(P)]$ . Ignoring the  $e^{-(1-\gamma)\varepsilon}$  (in fact, in appendix, we show that we can formally get rid of it by change of measure),  $E_t[U^L(P)] > U^L(E_t[P])$  by Jensen's inequality. As higher  $U^L$  represents lower utility, this shows that the convexity of value function creates more damage to L-type's utility upon crash than the expected reputation cost alone. Finally, if L-type chooses to take more risk, the loss of reputation always come at the same time as large financial loss. This aggravates the total loss in utility, which, given her risk-aversion  $\gamma > 1$ , is very undesirable.

As a result, L-type significantly reduces the level of latent risk-taking in contrast with the previous two cases. In equilibrium, she still chooses a higher level of latent risk over most reputation levels. But the gap between her level and H-type's level becomes smaller so that the expected difference in loss upon a crash is small relative to idiosyncratic jump loss  $\varepsilon$ . Consequently, crash loss cannot reveal a large amount of information about the manager's skill type. This leads to a "noisy pooling" equilibrium.

Suppose crash occurs when  $\hat{p}_{t-} = 0.5$ . Investors are most uncertain about the type of manager and updating is very sensitive to unexpected return shocks. The post-crash reputation  $\hat{p}_t = P(\hat{p}_{t-}, \phi_t^i, \varepsilon)$  depends on the realization of  $\varepsilon$ . We plot the probability density distribution of post-crash reputation  $\hat{p}_t$  for both types in Fig 1.7. The plot on the left panel shows the density distribution for both type of managers in an equilibrium with fully rational investors. H-type's density is shifted only slightly higher than that of L-type's, which suggests that investors cannot infer much from the crash loss. However, if managers assume that investors do not take their latent risk exposure into account and follow optimal strategy against naive investors, the information contained in the crash loss

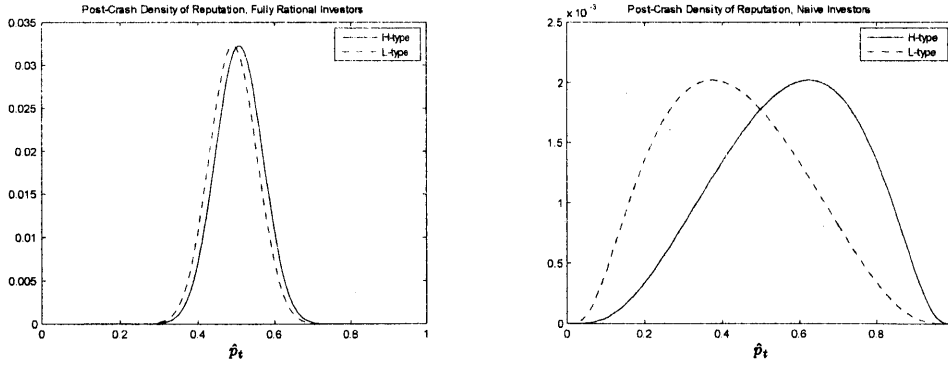
Figure 1-6: Component in L-type Value Function Due to Reputation



Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

is substantial. This is shown by the large difference in density distribution of post-crash reputation of H-type and L-type managers, which is plotted on the right panel of Fig 1.7. H-type's reputation improves considerably whereas L-type's reputation suffers.

Figure 1-7: Pdf of Post-Crash Reputation when  $\hat{p}_t = 0.5$



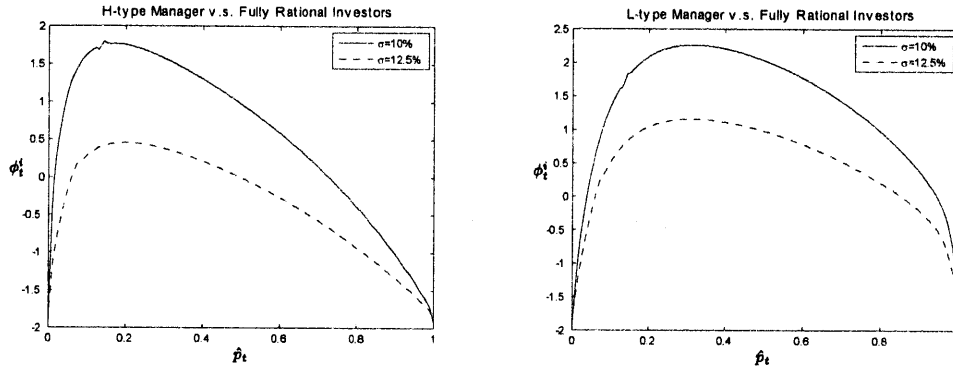
Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

Such a "noisy pooling" equilibrium suggests that the reputation change upon a crash would not be drastic. Rather, reputation change is gradual over time. It is often observed investors withdraw from funds after a fund suffers a large loss upon crash. It seems that one has to resort to other sources, such as fear of bank run or pure return chasing, to account for the large magnitude of withdrawal. Also, our model suggests that a manager who suffers more loss upon a crash may not necessarily be rationally downgraded. A skilled

H-type manager may put on more latent risk than L-type manager if her reputation is sufficiently low and investors' updating sensitivity is sufficiently high. This implies that, upon a crash, H-type loses more. Hence, investors should not lower a manager's reputation simply because of a higher level of crash loss when her reputation is already low.

Other features found in the "naive investors" case remain the same. As  $\alpha$  increases from 2% to 6%, we see the levels of optimal latent risk exposure are significantly higher for both types of managers in the fully rational equilibrium, which is shown in Fig 1.5. Greater information content in the realized fund return increases the sensitivity of updating by investors, which leads managers to load on more latent risk. Similarly, as shown in Fig 1.8, the levels of optimal latent risk exposure in fully rational equilibrium decrease for both types of managers when idiosyncratic risk  $\sigma$  increases from 10% to 12.5%. The reason is, again, the same as in the case with naive investors. More noise in the realized fund return decreases the sensitivity of updating by investors, which leads to lower latent risk taking.

Figure 1-8: Effect of Idiosyncratic Volatility (Against Fully Rational Investors)



Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  
 $\sigma_\epsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

## 1.6 Implications of the Equilibrium

In this section, we shall analyze the the implication of our equilibrium. We shall analyze the inefficiencies caused by the presence of latent risk and excessive risk taking. This is followed by other implications.

### 1.6.1 Inefficiencies

Excessive latent risk taking in equilibrium engenders several types of inefficiencies. Firstly, the choice of latent risk exposure taken by managers deviates from the risk-return optimal.

Surprisingly, even L-type managers suffer from latent risk exposure in equilibrium. Secondly, latent risk increases the time required to discover the skill level of a manager, which causes inefficiency in fund flow. Finally, we shall show that excessive latent risk exposure in equilibrium reduces the incentive for skilled managers to seek alpha opportunities, which aggravates the "slow-moving capital" problem.

### Distortion in Risk Taking

As shown in the analysis of fully rational equilibrium, both H-type and L-type managers take on excessive level of latent risk than what is optimal for pure investment purpose (i.e.  $\phi_t^i > \bar{\phi}$ ), which is damaging to their welfare. In equilibrium, L-type chooses to have a level of latent risk exposure that does not differ significantly from H-type so as to avoid being identified upon a crash. As a result, investors' conjecture of latent risk exposure taken by a manager is very close to the true exposure even though they do not know the exact type of manager. Consequently, they adjust their updating rule according to their conjecture and discount the portion of return that comes from latent risk. Therefore, excessive risk taking by both types fails to boost reputation in equilibrium relative to a situation where both types load on latent risk that is optimal from a pure investment perspective. But given a fixed investors' conjecture and the resulting updating rule, managers always have the incentive to boost reputation and cannot commit not to do so. Therefore, relative to a situation where both can commit not to take on excessive latent risk, managers' welfare are reduced significantly. This is true even for the unskilled ones at most reputation levels, who are supposed to benefit from the ability to mimic a skilled one through latent risk exposure in the short-term.

This problem echoes the call for third-party certification raised by some asset managers. Currently, SEC requires asset managers to disclose their holdings to public at quarterly basis. As discussed earlier, the frequency is too low and what is disclosed are mainly liquid and long positions, which are, furthermore, subject to the possibility of "window dressing". The complexity of financial contract as well as the large number of positions proves to be another source of obscurity. Hence, such disclosure does not provide enough information for investors to monitor the latent risk exposure  $\phi_t^i$  with precision to curb managers from loading on excessive level of latent risk. One may argue that the SEC could increase the frequency and the range of disclosed positions to allow better monitoring. However, this might discourage managers with true skill to spend effort and generate alpha if the alpha-generating strategy can be replicated easily going forward. Also, disclosure of holding position to public could lead to other serious problems such as predatory trading<sup>10</sup>. Thus, a third party is needed between managers and the investors to monitor the situation and provide certification to investors. If such a third-party is allowed for constant monitoring of manager's latent risk exposure and promises not to disclose anything, the above situation could be avoided. The third-party could make sure that manager's latent risk-taking is appropriate and certify to investors. And if investors classify managers who deviate from  $\bar{\phi}$  as L-type, then none of the types will deviate and take excessive latent risk. In this scenario,  $\phi_t^H = \phi_t^L = \bar{\phi}$  and investors will update without having to discount performance for latent risk-taking.

---

<sup>10</sup>See Brunnermeier and Pederson (2008), Chen, Hanson, Hong and Stein (2008).

We could conduct a welfare analysis based on H-type and L-type's value function. In particular, we shall follow Cole and Obstfeld (1991) and see how much AUM (in percentage) a manager of type  $i$  is willing to reduce in order to shift from the equilibrium with excessive latent risk exposure to the one with a third-party ensuring no excessive latent risk exposure. Under the equilibrium with excessive latent risk exposure, manager of type  $i$ 's future utility is given by her value function

$$V^i(W_t, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U^i(\hat{p}_t).$$

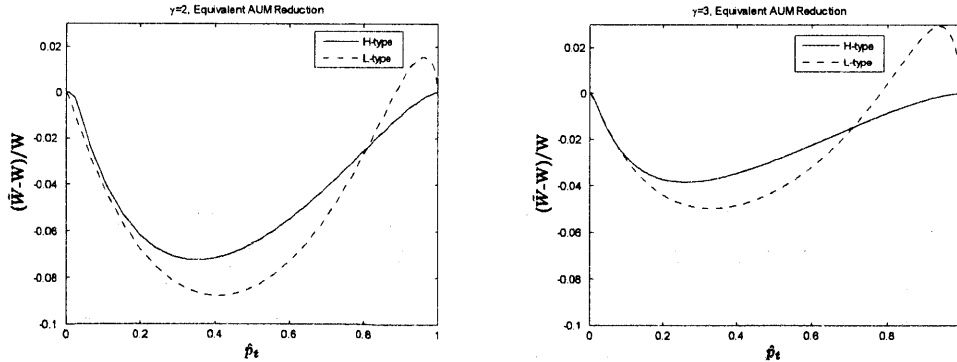
If managers takes no excessive level of latent risk exposure (i.e.  $\phi_t^i = \bar{\phi}$ ) and investors do not discount their performance, we can calculate the new value function

$$\tilde{V}^i(W_t, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} \tilde{U}^i(\hat{p}_t).$$

We find  $\tilde{W}_t$  such that  $\tilde{V}^i(\tilde{W}_t, \hat{p}_t, t) = V^i(W_t, \hat{p}_t, t)$  for current AUM  $W_t$  and reputation  $\hat{p}_t$ .  $\tilde{W}_t$  represents the equivalent AUM in an equilibrium with no excessive latent risk exposure that would deliver the same level of utility for the manager holding her current AUM and reputation fixed. We can compute the percentage reduction in AUM a manager of type  $i$  is willing to accept to avoid the excessive latent risk exposure and investors' discounting of her performance:

$$\frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U^i(\hat{p}_t) = \frac{1}{\rho} e^{-\rho t} \frac{\tilde{W}_t^{1-\gamma}}{1-\gamma} \tilde{U}^i(\hat{p}_t) \text{ implies } \frac{\tilde{W}_t - W_t}{W_t} = \left( \frac{U^i(\hat{p})}{\tilde{U}^i(\hat{p})} \right)^{\frac{1}{1-\gamma}} - 1.$$

Figure 1-9: Welfare Loss In Terms of AUM Reduction For Managers



Parameters Used:  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\alpha = 6\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  $\sigma_\epsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

For  $\gamma = 2$  and  $\gamma = 3$ , the equivalent AUM reduction for both types of managers at

different levels of reputation is plotted in Fig 1.9. We see the welfare loss due to excessive latent risk exposure is substantial for both types of managers with different levels of risk aversion. The loss is higher when risk aversion is low. This is because managers take on latent risk more aggressively to boost reputation when risk aversion is low. Also, the loss is higher for L-type than for H-type for most reputation levels as the former takes on more latent risk. The H-type is always willing to have third-party certification whereas L-type is willing to have when her reputation is not too high (which is unlikely).

In practice, such a third party must be diligent and trust-worthy. Otherwise, investors may not treat her certification seriously. Secondly, she must be sophisticated enough to detect latent risk exposure existing in a manager's portfolio. When latent risk is taken through certain assets well-known for negatively-skewed returns, detection is easy. However, if latent risk is taken through more complex forms of dynamic strategies, it requires considerable skill and effort from the third-party. But even if third-party monitoring is not perfect, it can greatly increase the cost of excessive latent risk exposure and ameliorate the problem. Besides, the presence of third-party monitoring could enhance proper enforcement of investment mandate that prohibits managers from taking latent risk ( $\phi_t^i = 0$ ), which might not be optimal but often better than the inefficient equilibrium. Most importantly, such a self-commitment implementation is actually in the interest of the fund managers themselves regardless of skill type. To a certain extent, our analysis explains the existence of fund administrators, who audits positions of the fund and ensures accurate pricing of some less frequently traded assets. Perhaps, more responsibilities should be assumed and more incentive for diligent monitoring should be provided.

We can also perform a similar analysis for investors, whose consumption rate is assumed to be proportional to the AUM. Of course, in reality, investors could leave a fund. Here, we consider a passive investor, who stays with the fund for simplicity. As a result, her wealth after withdrawal grows at

$$\frac{dW_t^I}{W_t^I} = (\mu^i + \pi\phi_t^i) dt + \sigma dB_t + [(1 - \Lambda\phi_t^i) e^{-L-\varepsilon} - 1] dN_t, \quad (1.25)$$

which differ from the manager's only in that there is no fund flow  $f(\hat{p})$  in the drift term. We assume that investors have the same utility function as the manager to remove the potential problems that arise from different attitudes towards risk. To a certain extent, investors choose to invest with funds that suit their risk appetite.

Conditioning on the manager, whom investors invest with, is of type  $i$ , investors have value function

$$\tilde{V}_{i,t}^I = \frac{1}{\rho} e^{-\rho t} \frac{(W_t^I)^{1-\gamma}}{1-\gamma} B_i^I \quad (1.26)$$

, where

$$B_i^I = \frac{\rho}{\rho + \lambda + (\gamma - 1) (\mu^i + \pi\bar{\phi} - \frac{1}{2}\gamma\sigma^2) - \frac{\pi}{\Lambda} (1 - \Lambda\bar{\phi})}$$

if manager takes "organic" level of latent risk exposure under third-party monitoring.

Thus, unconditionally, their value function will be

$$\tilde{V}_t^I(W_t^I, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{(W_t^I)^{1-\gamma}}{1-\gamma} \tilde{U}^I(\hat{p}_t) \quad (1.27)$$

where  $\tilde{U}^I(\hat{p}_t) = B_H^I \hat{p}_t + B_L^I (1 - \hat{p}_t)$ .

Without third-party monitoring, manager will take on excessive level of latent risk. Under managers' equilibrium strategy, we can still calculate investors' value functions conditioned on manager's type

$$V_{i,t}^I(W_t^I, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{(W_t^I)^{1-\gamma}}{1-\gamma} U_{i,t}^I(\hat{p}_t). \quad (1.28)$$

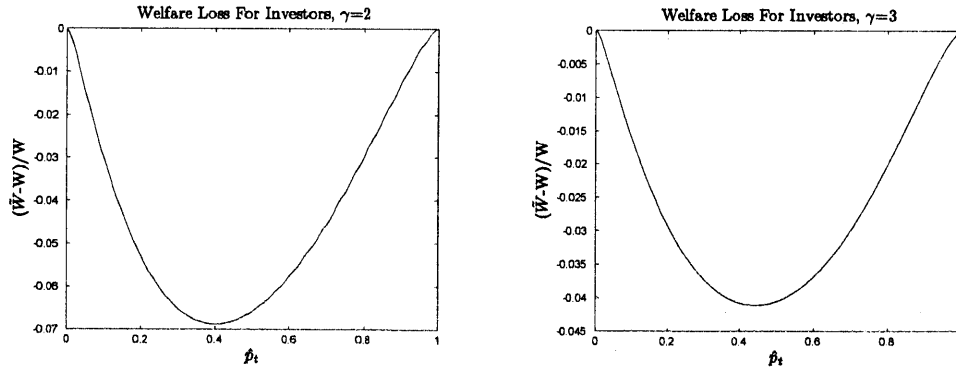
Notice that though the reputation of asset manager  $\hat{p}_t$  does not affect investors through fund flow, it affects manager's risk-taking, which still feeds back into investors' value function. Unconditionally, investors have value function

$$V_t^I(W_t^I, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{(W_t^I)^{1-\gamma}}{1-\gamma} U_t^I(\hat{p}_t) \quad (1.29)$$

, where

$$U_t^I(\hat{p}_t) = U_{H,t}^I(\hat{p}_t) \hat{p}_t + U_{L,t}^I(\hat{p}_t) (1 - \hat{p}_t).$$

Figure 1-10: Welfare Loss In Terms of AUM Reduction For Passive Investors



Parameters Used:  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\alpha = 6\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\% \hat{p}_t$

We can perform the similar analysis as we did for managers and compute the equivalent reduction in investment for passive investors without third-party monitoring. As shown in Fig 1.10, We see that substantial loss is incurred for investors especially when manager's



Table 1.1: Expected Discovery Time

$\alpha$	No Latent Risk		With Latent Risk	
	L-Type (Yrs)	H-Type(Yrs)	L-Type(Yrs)	H-Type(Yrs)
4%	28.3	28.3	40.3	42.7
5%	17.9	17.9	23.7	25.6
6%	12.5	12.5	16.5	17.7

Parameters Used:  $\gamma = 3$ ,  $\rho = 0.2$ ,  $\bar{\mu} = 10\%$ ,  $\sigma = 10\%$ ,  $L = 15\%$ ,  $\sigma_\varepsilon = 20\%$ ,  $\pi = 1\%$ ,  $\Lambda = 4.35\%$ ,  $\lambda = 0.2$ ,  $f(\hat{p}_t) = -12.5\% + 25\%\hat{p}$

risk aversion is low, which implies greater excessive risk-taking. Moreover, the loss is most severe when  $\hat{p}_t$  is in the middle when information asymmetry is most severe.

### Type Discovery

Here we shall study the existence of latent risk on discovery of manager's type by investors. If managers are inaccessible to latent risk, then difference of return drift is simply the true alpha difference between H-type and L-type managers. This is also the same if managers commit to,  $\bar{\phi}$ , the pure investment optimal through third-party monitoring. And this difference is what investors rely on to distinguish manager's skill. As more and more historical returns get observed, investors update on the type of manager they face and, given sufficiently long time, rational updating will converge to the true type<sup>11</sup>.

In the presence of latent risk, the difference in the expected return is no longer the true alpha. It contains the difference in latent risk exposure. For most of domain of reputation  $\hat{p}_t$ , L-type takes more latent-risk in equilibrium. This effect will reduce the sensitivity

$$\beta_{\text{With latent risk}} = \frac{\hat{p}_t(1 - \hat{p}_t)(\alpha + \pi\Delta)}{\sigma^2} < \frac{\hat{p}_t(1 - \hat{p}_t)\alpha}{\sigma^2} = \beta_{\text{No latent risk}} \text{ if } \Delta < 0. \quad (1.30)$$

In addition, the expected outperformance that H-type could generate,

$$E_t[dR_t|i = H] - E_t^I[dR_t] = (1 - \hat{p}_t)(\alpha + \pi\Delta) \quad (1.31)$$

, is also reduced if  $\Delta < 0$ . As a result, the magnitude of positive drift of  $\hat{p}_t$  for a H-type manager decreases with both sensitivity of updating and expected outperformance lowered in the presence of latent risk. This will slow down the speed of discovery and the capital movement from unskilled managers to skilled managers. Of course, another channel of updating is the crash loss. But as discussed earlier, since the difference in latent risk exposure taken by H-type and L-type managers is small relative to the idiosyncratic loss, there is limited information revealed by the crash loss, which is already rare in nature. Thus, latent risk will lengthen the process of discovery of managers' true skill levels by investors. In equilibrium, H-type's latent risk exposure could be slightly higher (i.e  $\Delta > 0$ ) when her reputation is sufficiently low. This would reverse the situation discussed above. But given that this happens only when  $\hat{p}_t$  is close to 0 and the sensitivity of updating

<sup>11</sup> $\lim_{t \rightarrow \infty} \hat{p}_t = 0$  or 1 almost surely.

is low, it offers little improvement in the speed of discovery. Thus, in the presence of latent risk, the speed of discovery is significantly lowered. In Table 1.1, we calculate the expected time for a new entrant H-type manager with  $\hat{p}_t = 0.5$  to reach  $\hat{p}_t = 0.9$  (investors are 90% sure of the manager's type) and a L-type manager with  $\hat{p}_t = 0.5$  to reach  $\hat{p}_t = 0.1$  (investors are 90% sure of her true type) through simulation. For  $\alpha = 4\%, 5\%, 6\%$ , we find that the expected discovery time is about 30% – 40% longer in an equilibrium with latent risk. Substantial inefficiency in fund flow is caused as a result.

### Slow-Moving Capital

It has been noted that arbitrage capital is slow-moving<sup>12</sup>, which leaves profitable investment opportunities unexploited for an extended period of time. Several papers have attempted to provide explanations behind this puzzling phenomenon. For instance, He and Xiong (2010) argues that optimal contracts may be designed to elicit effort from managers by restricting the set of investment opportunities, which leads to slow capital movement. Malliaris and Yan (2010) argues that managers, who suffers reputation damage, may forgo profitable opportunities for "nickel-picking" strategies. Our results suggest that the presence of latent risk might aggravate the problem.

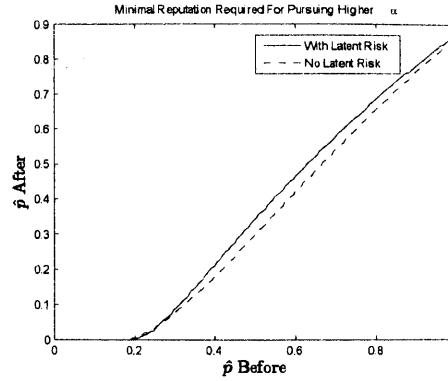
When new profitable opportunities arise, it can be exploited by skilled managers. If truly skilled manager pursues a new strategy with higher alpha  $\alpha' > \alpha$ , there might be a fixed cost associated with such a move. Thus, only if a significant improvement in  $\alpha$  could be attained, would a skilled manager be willing to pay the cost and pursue the new opportunities. Higher level of alpha increases the sensitivity of investors' updating rule, which induces higher latent risk exposures. As discussed earlier, this only causes further deviation from the pure investment optimal without having much actual performance boosting effect since, in equilibrium, investors would expect a higher level of performance manipulation through latent risk and discount more from the observed fund return. Further deviation from the pure investment optimal  $\bar{\phi}$  leads to lower marginal benefit of higher alpha. With lower marginal benefit, skilled manager requires larger improvement  $\alpha' - \alpha$  to compensate for the cost of capital movement. She would forgo more alpha improvement opportunities, which aggravates the slow-moving capital problem.

For instance, in the context of our model, a fixed cost could manifest as a reduction in reputation, if investors are not sure about a manager's ability to succeed with a new trading style. Holding fixed the potential increase from  $\alpha$  to a higher level  $\alpha'$  brought about by the new profitable opportunities, a skilled manager will only decide to exploit the opportunities if the decrease in  $\hat{p}_t$  is not too large. Fig 1.11 provides a numerical illustration for the case, in which pursuing new opportunities will increase  $\alpha$  from 2% to 6%. The dashed line represents the minimal level of reputation (after the move) for a skilled H-type manager to be willing to implement such a move at different levels of reputations (before the move) when there is no latent risk available. In the presence of latent risk, the situation is aggravated due to lower marginal benefit of higher alpha as discussed above. This in general leads to higher level of minimal reputation (the solid line

---

<sup>12</sup>It has been noted in Duffie (2010). Mitchell, Pedersen and Pulvino (2007) provides an example by studying the convertible arbitrage funds.

Figure 1-11: Minimal Level of Reputation Required



Parameters Used:  $\rho = 0.2, \alpha = 2\%, \alpha = 6\%, \bar{\mu} = 10\%, \sigma = 10\%, L = 15\%$   
 $\sigma_\epsilon = 20\%, \pi = 1\%, \Lambda = 4.35\%, \lambda = 0.2, f(\hat{p}_t) = -12.5\% + 25\%\hat{p}_t$

in Fig 1.11) required for a skilled H-type manager. Note that the exact form of fixed cost is not important here. As long as the excessive latent risk exposure in equilibrium lowers the marginal benefit of higher alpha, we would expect an aggravation of the "slow-moving capital" problem when there is an associated fixed cost for capital movement.

## 1.6.2 Other Implications

### Subprime Mortgage Crisis

Our study has interesting implications on the subprime mortgage crisis. As argued earlier on, mortgage-backed securities generate good returns when market is under normal conditions but incur large loss upon market or liquidity crash. This feature fits the nature of latent risk. Furthermore, subprime mortgage securities are illiquid, which cannot be marked to market. Thus, it generates little Gaussian risk. As many of these products were highly rated, asset managers were not restricted from holding it. Thus, latent risk became readily available for a broad class of asset managers ranging from the conservative pension fund managers to the aggressive hedge fund managers.

During the period of 2002 to 2006, market was awash with liquidity. Existing funds had seen AUM growing rapidly and many new funds were set up. This is best illustrated by the growth of hedge fund industry from around 0.6 trillion to around 2 trillion before the crisis. As our model suggests, asset managers in general take on positive amount of latent risk unless their skill is clearly recognized with reputation  $\hat{p}_t$  close to 0 or 1. With more asset under management, this creates a larger demand for latent risk in equilibrium as demonstrated earlier on. Moreover, with continuous securitization effort made by mortgage originators and investment banks, the supply of latent risk is also increasing, which ensures the attractiveness of latent risk premium to asset managers even in the pres-

ence of greater demand. This partially explains the excessive demand and over-supply of subprime mortgage-backed securities.

Moreover, given the supply of skilled managers is limited, the overall proportion of managers with skill to generate alpha is reduced. This reduces the prior belief of manager being skilled held by investors. As our equilibrium suggests that the peak of latent risk taking is very much skewed towards the left (see Fig 1.5)<sup>13</sup>. So a decrease in ex-ante reputation of new entrants in general will cause them to load more on latent risk to boost reputation, which further increases the demand. Finally, as our model suggests for most of domain of reputation  $\hat{p}_t$ , the unskilled L-type takes on more latent risk than the skilled H-type. An increase in the overall proportion of unskilled managers also contribute to the increase in the overall demand for latent risk.

In summary, with the latent risk nature of mortgage-backed securities and easy accessibility to a broad range of asset managers, the increase of asset under management in the asset management industry led to greater demand for latent risk. This was accompanied by a rise in the proportion of unskilled managers and deteriorating prior reputation, which further stimulated the demand for latent risk and exacerbated the loss suffered by various fund styles during the crisis.

### Declining Return For Bigger Funds

It has been documented, for both mutual funds and hedge funds, return deteriorates with fund size<sup>14</sup>. This has been largely attributed to decreasing return to scale as a result of trading cost, limited arbitrage opportunities or organizational diseconomies etc. Our study complements these explanations with a novel argument. Bigger funds most likely have been successful in past performance. Regardless of managers' true skill type, good past performance led to high reputation. Our study suggests that asset managers with higher reputation  $\hat{p}_t$  for most of the domain of  $\hat{p}$  takes less latent risk. Thus, funds that enjoy high reputation  $\hat{p}_t$  will have lower return drift yielded from latent risk-taking if crash does not occur. Moreover, since latent risk should carry an insurance premium, the overall return after taking into account of crash loss is still positive. This provides a complementary explanation for declining return for bigger funds.

## 1.7 Fund Termination

In this section, we shall investigate features of managers' latent risk taking strategies when a shut-down condition is introduced. The motivation behind this is that, in reality, a manager may not be able to continue her career if her reputation becomes sufficiently low. Investors might withdraw money, which leads to fund termination. This could be because the investors are coordinated by the the reputation level and a low level of

---

<sup>13</sup>With fund termination, latent risk exposure becomes monotonically decreasing in reputation level. See next section.

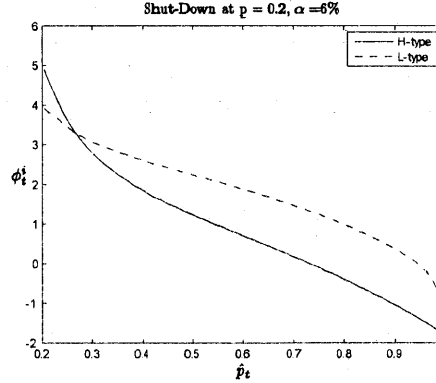
<sup>14</sup>E.g. Chen, Hong, Huang, Kubik (2004), Getmansky (2004)

reputation triggers a bank-run on the fund's capital. Alternatively, investors may find it more cost-saving to switch to passively managed exchange-traded-funds if the probability of a manager having alpha-generating skills is sufficiently low. If the manager works for a fund family, she might be fired for low reputation. In our model, we shall assume that the fund will be shut down if manager's reputation drops to a certain threshold level  $\underline{p} > 0$ <sup>15</sup>. We further assume that the manager's continuation utility after fund termination is of the form

$$\frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} \underline{U}^i$$

with  $\underline{U}^i = KU^i(\underline{p})$  for some  $K > 1$ . The functional form of the continuation utility is chosen to be similar to the value function of the manager for tractability purpose. The chosen functional form implies that the shut-down continuation utility is higher for a manager with higher AUM. This could be interpreted in several ways. The manager might have a stake in the fund, which would be returned to her upon fund termination. Alternatively, a manager with experience managing a larger fund will in general be able to manage more asset should there be an opportunity for her to start a new career elsewhere. As discussed earlier on, higher  $U^i$  represents lower utility. By imposing  $K > 1$ , we assume the fund manager suffers a loss from the termination. For a numerical illustration, we set  $\underline{p} = 0.2$  and  $K = 1.1$ .

Figure 1-12: Minimal Level of Reputation Required



Parameters Used:  $\gamma = 3, \rho = 0.2, \alpha = 2\%, \bar{\mu} = 10\%, \sigma = 10\%, L = 15\%$   
 $\sigma_\epsilon = 20\%, \pi = 1\%, \Lambda = 4.35\%, \lambda = 0.2, f(\hat{p}_t) = -12.5\% + 25\%\hat{p}_t$

The level of latent risk taking is demonstrated in Fig 1.12. Latent risk exposure becomes monotonically decreasing in reputation  $\hat{p}_t$ . Thus, the inverse U-Shape might not hold if we introduce forced termination. As expected, H-type becomes more and more aggressive in risk-taking as  $\hat{p}_t$  approaches 0.2, which reflects her eagerness to avoid termination. The cost for her is much higher than an unskilled manager given her alpha-generating ability and associated potential to attract fund inflow in the long run. Apart from the inverse

<sup>15</sup>See Malliaris and Yan (2010) for similar assumptions

U-Shape, the other properties remains intact. Both types of managers still take excessive amount of latent risk. L-type has higher exposure except when the reputation level  $\hat{p}_t$  is sufficiently low. And the equilibrium difference in levels of latent risk exposure taken by H-type and L-type is small, which can be regarded as "noisy pooling".

## 1.8 Empirical Analysis

Latent risk exposure provides small steady return under normal market conditions but incurs large crash loss. Its risk-return profile naturally leads to crash loss if a manager decides to load on it to boost performance. It has been well-documented that hedge fund returns exhibit negative skewness due to rare extreme losses. Among all, the most famous case is what happened to LTCM in 1998 after the default of Russian government bond. The collapse of the 9 billion hedge fund Amaranth in 2006 due to an extreme loss of more than 65% provides another example. Though not as dramatic, hedge funds returns are in general negatively-skewed. And, given the secrecy of their portfolio positions and flexibility of their investment mandates, hedge funds managers have easy access to latent risk and are, perhaps, most likely to boost reputation through latent risk exposure. Thus, we shall focus our empirical tests on hedge funds, although as discussed earlier on, performance manipulation through latent risk exposure is not restricted to the hedge fund industry alone.

Table 1.2: Dow Jones Credit Suisse Indices Return

Indices	Sector Weight	Mean(%)	Stdev(%)	Skewness
Overall	-	0.78%	2.22%	-0.22
Convertible Arb	3.0%	0.65%	2.04%	-2.76
Dedicated Short	0.5%	-0.20%	4.92%	0.70
Emerging Markets	9.8%	0.76%	4.39%	-0.78
Equity Neutral	3.7%	0.48%	3.07%	-11.79
Event Driven	21.9%	0.84%	1.76%	-2.44
Fixed Income Arb	1.9%	0.44%	1.71%	-4.32
Global Macro	10.5%	1.03%	2.90%	-0.03
Long/Short	23.3%	0.86%	2.88%	-0.01
Managed Futures	13.8%	0.59%	3.41%	0.00
Multi Strategy	11.7%	0.66%	1.56%	-1.76

Table 1.2: Data consist of monthly returns of the Dow Jones Credit Suisse indices starting from January 1994 until December 2010. The indices tracks about 8000 funds with minimum 50 million AUM. The index returns are weighted by AUM and net of fee.

Table 1.2 documents the mean, standard deviation and normalized skewness of the Dow Jones Credit Suisse (formly known as the Tremont indices) hedge fund indices returns. Most of fund style categories exhibits negative skewness with the exception of managed

futures and dedicated short bias, the latter of which has negligible sector weight. The overall index, which tracks all hedge funds performance, has negative skewness of  $-0.22$ , which is suggestive of the general trait of hedge fund return.

While returns of different hedge fund styles seem to be negatively skewed, one cannot link this feature to the potential latent risk exposure directly. Hedge funds load on well-known risk factors and the negative skewness could come from these factor exposures. Performance variations related to well-known observable factors should be filtered off by investors when evaluating a manager. Exposures on well-known factors do not help a manager boost performance and should not be considered as latent risk exposure. Thus, latent risk should be contained in the idiosyncratic component of fund returns that cannot be explained by loadings on factors. As our model is about endogenous choice of latent risk, we shall work on idiosyncratic component of the fund returns after removing factor exposures.

We shall employ the seven-factor model proposed by Fung and Hsieh (2004) as the benchmark for calculating idiosyncratic component of fund returns. Two factors are equity-oriented: excess return on the CRSP value-weighted market portfolio (MKT) of NYSE, AMEX and NASDAQ stocks as well as the Fama-French size factor (SMB). Two factors are fixed-income driven: change in the constant-maturity yield on U.S. 10-year Treasury bond (YLDCHG) as well as change in credit spread (CRDSPD) between yields of Moody's Baa corporate bond and U.S. 10-year Treasury bond. Finally, three trend-following strategy factors<sup>16</sup> are included: the return of bond lookback straddles (TFBD), the return of currency lookback straddles (TFFX) and the return of commodity lookback straddles (TFCM). An examination of the factors shows that many of them indeed have substantially skewed returns. This is documented in Table 1.3. Thus, depending on the signs of loadings, these risk factors could contribute to the skewness of fund returns positively or negatively.

To single out the idiosyncratic component, in which latent risk exposure could possibly play a role, we shall regress the hedge fund returns on the risk factors to obtain the residuals. We focus on those idiosyncratic residuals large in absolute value as they are more likely to be related to the potential latent risk exposure where as the small residuals are largely produced by Gaussian shocks. Our model generates several interesting predictions. Firstly, the higher the volatility of the idiosyncratic component, the lower the level of latent risk exposure. This is because higher volatility of the idiosyncratic component leads to lower sensitivity of investors' updating, which reduces the incentive for managers to boost return through latent risk exposures. This is illustrated by Fig 1.8, in which higher  $\sigma$  leads to lower  $\phi_t^i$  at all reputation levels. Therefore, large idiosyncratic movements should be more positive if idiosyncratic volatility is higher. This is **prediction 1**.

---

<sup>16</sup>Trend-following factor data are kindly made available on David Hsieh's website <http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm>. See Fung and Hsieh (2001) for a description of these factors.

Table 1.3: Hedge Fund Benchmark Factors

Factors	Mean	Stdev	Skewness
Market	0.47%	4.65%	-0.86
SMB	0.21%	3.72%	0.87
YLDCHG	-0.01%	0.29%	-0.10
CRDSPD	1.17%	0.86%	2.05
TFBD	-1.64%	14.78%	1.46
TFFX	0.02%	19.72%	1.40
TFCM	-0.47%	13.89%	1.29

Table 1.3: Data consist of monthly return from January 1994 to March 2010 of 7 risk factors employed in Fung and Hsieh (2004). The seven factors are: excess return of CRSP value-weighted market portfolio, Fama-French size factor, constant-maturity yield change on U.S. 10-year Treasury bond, change in credit spread between Barcap U.S. Corporate bond BAA and AAA indices, the return of bond lookback straddles, the return of currency lookback straddles and the return of commodity lookback straddles.

Secondly, the higher the difference in return generated by skilled and unskilled managers, the higher the level of latent risk exposure and the higher the loss upon a crash. This is again due to sensitivity of investors' updating. In our model, the differential return is captured by parameter  $\alpha$ . Higher  $\alpha$  suggests higher information content relative to noise in the unexpected return shocks perceived by investors, which increases the sensitivity of updating and induces higher latent risk exposure. This is illustrated by Fig 1.1, Fig 1.4 and Fig 1.5 for various levels of investors' rationality. This suggests that large idiosyncratic movements should be decreasing in the difference in alpha-generating ability between the skilled and unskilled managers. This is **prediction 2**.

Finally, our model predicts the level of latent risk exposure is related to the reputation level of the manager. Without fund termination, we expect that as manager's type becomes more clear to the investors ( $\hat{p}_i$  near 0 or 1), they would take less latent risk exposure. As discussed before, when investors are relatively certain of managers' type, they do not update aggressively and this lowers the incentive for managers to take on latent risk. With fund termination, managers with sufficiently low reputation are driven out. Reputation level is not directly observable. Calculated measures suffer two major problems. There is no single unanimous measure that summarizes the reputation perceived by investors. A variety of methods have been proposed. Jensen's Alpha and Sharpe ratio are the best known ones. But among hedge funds, maximum draw-down is also a key statistic. Also, in practice, many investors judge the manager by looking at how much her absolute return beats certain benchmark. Besides, Bayesian approach and measures based on portfolio holdings shown up in 13-F filings have also been suggested. The second problem is that any measure of reputation is, by its purpose, is a measure of skill. In addition to alpha-generating ability, a skilled manager is likely to be good at risk management, which we do not include in our model. Better risk management reduces loss upon crash. This will interfere with the relation between crash loss and reputation level. We propose a



solution that uses fund age and asset under management (AUM) as a proxy. As a fund's age increases, more historical performance becomes available for investors to update on manager's skill type and a manager's type will be revealed eventually. This is shown by the fact that, in equilibrium, the reputation level  $\hat{p}_t$  tends to 0 or 1. Skilled managers will lower latent risk exposure as the fund age increases. Unskilled managers will increase latent risk exposure in the beginning to boost performance as their reputation declines over time. As age further increases, funds run by unskilled managers will either be terminated or reduce latent risk exposure as investors have pinned down their type and no longer update. Since skilled managers are scarce, the effect of unskilled manager dominates in the beginning. Thus, large idiosyncratic movements should decrease in age first. However, as funds run by unskilled managers are terminated or detected with little uncertainty, large idiosyncratic movements should start to increase in age. Therefore, we expect U-shaped relation between large idiosyncratic movements and fund age. AUM is another proxy. Large fund tends to do well in history and enjoys higher reputation regardless of the true skill level of the manager. This is **prediction 3**.

### 1.8.1 Data

We obtain hedge fund returns, age and other fund-specific characteristics from the Lipper TASS database. There are "Live" funds and "Graveyard" funds. The "Live" ones are active until the last update of the TASS database, which was March 2010 in our data. "Graveyard" funds are those that stopped reporting to the database before March 2010 due to various reasons such as liquidation or closed to new investors. "Graveyard" funds were created in 1994 to mitigate "survivorship" bias. As a result, we exclude returns reported to the database prior to 1994.

We only include funds that are reporting returns net of fee in US dollars at monthly frequency. This covers the majority of the funds in the database. Besides, an additional condition on fund size is imposed. We require the average fund size over its history to be greater than 10 million. Similar requirement is imposed in Hu, Pan and Wang (2010) and Cao, Chen, Liang and Lo (2010) studying the same dataset<sup>17</sup>. To ensure proper estimation of loadings on risk factors, we exclude those funds that have less than 36 observations (3 years). The summary statistics for 11 fund styles are presented in Table 1.4.

### 1.8.2 Idiosyncratic Component

To recover the idiosyncratic component of fund returns, we can regress individual fund returns on the 7 factors specified above to remove systematic risks. Given the large autocorrelation of hedge fund returns<sup>18</sup>, we shall include the lagged market return in the regressors<sup>19</sup> as well.

---

<sup>17</sup>Other cutoff criterion are also experimented with and empirical inferences remain largely unaffected.

<sup>18</sup>e.g. Getmansky, Lo and Makarov (2004)

<sup>19</sup>We experiment without lagged market return and our results are unaffected.

Table 1.4: TASS Data Summary Statistics

	No. of Funds		Avg Ret(%)		Stdev(%)		Skewness		Fund Size(\$M)		Age		No. of Obs	
	Total	Graveyard	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med
All	5553	2796	0.59	0.53	3.02	2.45	-0.68	-0.53	334	60	3.8	3.1	85	73
Convertible Arb	119	87	0.64	0.61	2.44	1.86	-1.23	-0.74	153	79	4.4	3.6	96	85
Emerging Markets	244	100	1.01	0.92	4.83	4.63	-0.75	-0.44	137	52	3.6	3.1	83	72
Market Neutral	172	121	0.58	0.53	2.49	2.10	-0.33	-0.08	94	44	3.5	3.0	81	71
Event Driven	355	234	0.80	0.77	2.50	2.13	-0.65	-0.66	203	75	4.5	3.5	96	84
Fixed Income Arb	132	100	0.61	0.61	2.35	2.05	-1.66	-0.97	220	95	3.8	3.5	89	79
Fund of Funds	1175	596	0.47	0.45	2.32	2.01	-0.93	-0.82	170	61	4.1	3.4	91	81
Global Macro	141	80	0.73	0.72	3.59	3.20	0.16	0.19	259	59	3.9	3.0	80	67
Long/Short Equity	1029	603	0.88	0.85	4.19	3.79	0.02	-0.03	119	51	4.1	3.3	90	77
Managed Futures	177	72	0.76	0.78	4.33	4.14	0.22	0.24	202	51	5.2	4.0	103	85
Multi-Strategy	204	112	0.74	0.68	2.89	2.52	-0.58	-0.53	231	68	3.7	2.8	84	67
Undefined	1776	677	0.35	0.28	2.62	2.12	-1.03	-0.88	699	62	3.1	2.8	74	66

Table 1.4: Summary statistics are calculated for all funds and funds in each fund category. "Avg Ret" and "Stdev" stand for average fund return and fund return standard deviation. "Skewness" stands for skewness of fund return. "Size" is the average AUM (Asset Under Management) of a fund over the time span of report. "Age" is the number of years between fund's inception date and the date of return report. If inception date is not available, we use the date of the first report. "No. of Obs" stands for number of monthly observations for a fund.

Thus, our regression is specified as

$$\begin{aligned}
r_{i,t} = & \alpha_i + \beta_{MKT}MKT_t + \beta_{Lag\_MKT}MKT_{t-1} + \beta_{SMB}SMB_t + \beta_{YLDCHG}YLDCHG_t \\
& + \beta_{CRDSPD}CRDSPD_t + \beta_{TFBD}TFBD_t + \beta_{TFFX}TFFX_t \\
& + \beta_{TFCM}TFCM_t + \varepsilon_{i,t}
\end{aligned} \tag{1.32}$$

where  $r_{i,t}$  is the excess return of fund  $i$  in month  $t$  and  $\varepsilon_{i,t}$  is the idiosyncratic component of fund  $i$  in month  $t$ .

As it is well-known that hedge fund risk exposures are time-varying<sup>20</sup>, we shall resort to the rolling-window regressions with window size of 36 months. This allows us to compute residuals  $\hat{\varepsilon}_{i,t}$  and the volatility  $\hat{\sigma}_i^\varepsilon$  of the residuals  $\hat{\varepsilon}_{i,t}$ . The results are summarized in Table 1.5. We can calculate the volatility and skewness of the residual errors from the regression, which measure the volatility and skewness of the idiosyncratic component of the fund return. The skewness of the residual errors are much less negative than the skewness of fund returns, which is evidenced by a drop in average skewness from -0.68 to -0.14. This suggests that a large portion of the negative skewness of hedge fund returns could be attributed to loadings on risk factors that are skewed. However, the average skewness is still negative for most of the fund categories. This alludes to the potential latent risk taking behavior studied in this paper.

Table 1.5: Seven Factor(plus Lagged Mkt) Regression

	Alpha(%)		Idiosyncratic Vol		Idiosyncratic Skewness	
	mean	med	mean	med	mean	med
All	0.32	0.24	1.96	1.59	-0.14	-0.09
Convertible Arb	0.38	0.35	1.48	1.33	-0.26	-0.29
Emerging Markets	0.68	0.66	3.10	3.06	0.00	0.10
Market Neutral	0.37	0.31	1.85	1.64	-0.27	-0.02
Event Driven	0.53	0.48	1.60	1.32	0.04	0.03
Fixed Income Arb	0.45	0.38	1.64	1.36	-0.78	-0.33
Fund of Funds	0.17	0.15	1.42	1.18	-0.28	-0.21
Global Macro	0.39	0.41	2.62	2.36	0.11	0.10
Long/Short Equity	0.54	0.48	2.72	2.49	0.16	0.11
Managed Futures	0.41	0.39	3.18	3.00	-0.07	-0.04
Multi-Strategy	0.52	0.40	1.96	1.74	-0.03	-0.01
Undefined	0.12	0.05	1.67	1.35	-0.27	-0.19

Table 1.5: Individual fund return is regressed on 7 hedge fund risk factors and lagged market returns. Alpha, standard deviation and skewness of residual errors of fund-level regressions are reported.

<sup>20</sup>e.g. Bollen and Whaley (2009)

### 1.8.3 Testing The Predictions

As managers could have different risk appetites or strategies, funds have different levels of idiosyncratic volatility. We shall rescale residuals by dividing the volatility  $\hat{\sigma}_i^\varepsilon$  to make it more comparable across funds. The scaled residual is  $\hat{\xi}_{i,t} = \hat{\varepsilon}_{i,t}/\hat{\sigma}_i^\varepsilon$ . Moreover, as we are only interested in latent risk that will incur relatively large realized idiosyncratic component, we shall select large scaled residuals –  $\hat{\xi}_{i,t}$ ’s with absolute values larger than 3<sup>21</sup>, which could happen with very small chance for Gaussian noise. This will avoid sample dilution by small idiosyncratic movements caused largely by Gaussian shocks, which is unrelated to latent risk exposure.

We shall test our predictions by regressing the large rescaled residuals on several predictors. Prediction 1 suggests that higher residual volatility  $\hat{\sigma}_i^\varepsilon$  should lead to higher  $\hat{\xi}_{i,t}$ . Prediction 2 requires the difference in ability between skilled and unskilled managers perceived by the investors as predictor. As this is not directly observable, we shall infer it from the data. Specifically, for fund style  $j$  listed in Table 1.4, we consider the empirical distribution of realized Sharpe ratio 6 months ago. Each fund  $i$  that belongs to style  $j$  has a realized excess return 6 months ago  $R_{i,t-6}$ . We divide it by the volatility of the fund to obtain the realized Sharpe ratio of fund  $i$ . Of course, this would be a noisy measure as it is based on only one observation. However, the empirical distribution of realized Sharpe ratio would be quite stable by law of large number. Thus, we can measure,  $AD_{j,t}$ , the ability difference between skilled and unskilled managers in style  $j$  by the inter-quartile range of the empirical distribution and assign it each fund  $i$  in style  $j$ . We use the empirical distribution 6 months ago because it would take time for a manager to adjust her latent risk exposure. The results are qualitatively the same if we use a lag of 3 months. Prediction 2 suggests higher  $AD_{i,t}$  should lead to lower  $\hat{\xi}_{i,t}$ . Finally, prediction 3 indicates that there is a U-shape relationship between  $\hat{\xi}_{i,t}$  and age as well as a positive relationship between  $\hat{\xi}_{i,t}$  and AUM. To capture the non-monotonic relationship, we impose a quadratic functional form and use  $Age_{i,t}$  and  $Age_{i,t}^2$  as two additional predictors. For AUM, we shall follow the common practice and take the logged value. In total, there are 5 predictors: residual volatility  $\hat{\sigma}_i^\varepsilon$ , ability difference perceived by investors  $AD_{i,t}$ ,  $Age_{i,t}$ ,  $Age_{i,t}^2$  as well as  $\log(AUM_t)$ .

Specifically, the regression is as follows

$$\begin{aligned} \hat{\xi}_{i,t} = & c + \phi_1 \hat{\sigma}_i^\varepsilon + \phi_2 AD_{i,t} + \phi_3 Age_{i,t} + \phi_4 Age_{i,t}^2 + \phi_5 \log(AUM_t) \\ & + \phi_6 \text{Ind}_{\{\text{Fund Style}\}} + \phi_7 \text{Ind}_{\{\text{Year}\}} + \phi_8 X_{i,t} + \eta_{i,t} \end{aligned} \quad (1.33)$$

where  $\text{Ind}_{\{\text{Fund Style}\}}$  and  $\text{Ind}_{\{\text{Year}\}}$  are dummy variables for fund styles and years and  $X_{i,t}$  are controls. Dummies for fund styles and years are also included to remove the potential unobserved style fixed effects and changing economic conditions over time. Percentage below high-water mark<sup>22</sup> and contemporaneous Dow Jones Credit Suisse style return are

<sup>21</sup>Different cutoff thresholds (2 and 4) are also examined.

<sup>22</sup>Percentage below high-water mark is defined as (Current NAV per share - Maximum NAV attained in history per share)/Maximum NAV attained in history per share.

included as the main controls. The contemporaneous style return is important because it controls for the impact of economic fundamentals on a typical fund of a style. We expect  $\phi_1 > 0$  since residual volatility will reduce latent risk exposure. Furthermore, we expect  $\phi_2 < 0$  as higher expected difference in ability between skilled and unskilled managers will induce higher latent risk exposure. Since latent risk exposure increases with age initially, we expect  $\phi_3 < 0$ . After funds run by unskilled managers are terminated or identified with little uncertainty, the overall effect of age on latent risk exposure will flip sign and we expect  $\phi_4 > 0$ . Finally, as large funds tend to have higher reputation, which reduces latent risk exposure, we expect  $\phi_5 > 0$ .

Table 1.6 summarizes the findings. In specification (1), we regress large rescaled residuals computed from the 7-factors regression (plus lagged market excess return) for each individual fund on the standard deviation of the residuals. We get a positive and statistically significant relation. This indicates that funds with more idiosyncratic volatility tend to have smaller loss. This is consistent with our model prediction in that when idiosyncratic volatility is high, investors update less aggressively to unexpected good performance, which reduces manager's incentive to load on latent risk. In specification (2), we regress large scaled residual on  $AD_{i,t}$ , the difference in alpha-generating ability between skilled and unskilled managers of the style that the fund belongs to. We obtain a negative and significant relation. This suggests that when investors expect skilled managers to generate high alpha, managers increase latent risk exposure to boost performance. In specification (3), we test the relationship between large scaled residual and fund age as well as fund size, both of which are proxies for the unobserved reputation. As expected, we obtain a negative coefficient on age and a positive coefficient on squared age, which are both statistically significant. This generates a quadratic U-shape relationship and confirms the prediction that latent risk exposure is increasing in age initially when the unskilled managers dominate the population but is decreasing in age as time goes by and unskilled managers are driven out or identified. The coefficient on logged AUM is also positive and significant, suggesting that large funds tend to suffer smaller loss. In specification (4), we put idiosyncratic volatility, difference in alpha-generating ability and age effects together on the right hand side of the regression. Our results remain robust. All the coefficients are still significant. The magnitudes of coefficients are stable across 4 specifications as well.

#### 1.8.4 Robustness Checks and Discussion

One concern is the choice of cutoff threshold. We experiment with different cutoff thresholds for large residuals. Table 1.7 reports the results for regression (1.27) using threshold = 2, 3, 4. For all 3 cases, the mean of large residuals is negative and more so when the cutoff is higher. This indicates that large movements are more likely to be negative, which is consistent with the negative skewness of residuals computed in Table 1.5 and suggestive of the presence of latent risk. The coefficients remain consistent with model predictions and statistically significant except for the coefficient on logged AUM with cutoff equal to 2. It is marginally insignificant perhaps due to sample dilution. As

Table 1.6: Large Residual Tests( > 3 Residual Vol)				
	Dependent Variable: Scaled Residual			
	(1)	(2)	(3)	(4)
Residual Vol	0.501 [5.59]			0.436 [4.97]
Ability Difference		-1.38 [-6.70]		-1.39 [-6.53]
Age			-0.415 [-5.64]	-0.409 [-5.68]
Age <sup>2</sup>			0.016 [4.17]	0.016 [4.42]
log(AUM)			0.436 [2.03]	0.463 [2.16]
Controls	Yes	Yes	Yes	Yes
Style Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Adj-R <sup>2</sup>	0.14	0.13	0.15	0.16
No. of Funds	3146	3146	3146	3146
No. of Obs	4676	4676	4676	4676

Table 1.6: Cross-sectional regression on large scaled residual are conducted under various specifications.

the threshold increases, the economic significance increases as evidenced by larger coefficients' magnitudes and higher adjusted  $R^2$ . This is expected because the observations are more likely to be induced by latent risk exposure. The turning point for the age effect, which is  $-\frac{\phi_3}{2\phi_4}$ , is consistently around 10 to 15 years.

Another concern is whether our results are driven by "live" funds or "graveyards" funds alone. We shall perform the same regression on the two sub-samples of "live" funds and "graveyard" funds separately. Results are presented in Table 1.8. The number of observations are roughly equal for the two subsamples. The sign of coefficients remain consistent and significant with model predictions.

For each individual predictor, we may come up alternative explanation for its significance. For instance, if, for reasons unrelated to reputation concern, a manager allocates more portfolio weights to assets or strategies exhibiting idiosyncratic Gaussian risk, she would have less portfolio weights on latent risk assets. This leads to more positive scaled residuals and higher residual volatility at the same time. Also, by including the style return and the year fixed effects as controls, we can largely remove the effects of time-varying investment opportunities driven by economic fundamentals. This might not perfectly eliminate the correlation with economic fundamentals, which would lead to spurious significance of the coefficient on  $AD_{i,t}$ . Moreover, fund with more symmetric idiosyncratic risk might have moderately large scaled residuals with equal probability of being positive or negative quite

Table 1.7: Different Cutoffs For Large Residuals

	Dependent Variable: Scaled Residual		
	> 2 Res Vol	> 3 Res Vol	> 4 Res Vol
Residual Vol	0.133 [4.46]	0.436 [4.97]	1.27 [5.27]
Ability Difference	-1.14 [-13.90]	-1.39 [-6.53]	-1.58 [-3.03]
Age	-0.155 [-7.23]	-0.409 [-5.68]	-0.662 [-3.68]
Age <sup>2</sup>	0.007 [6.56]	0.016 [4.42]	0.022 [2.17]
log(AUM)	0.121 [1.81]	0.463 [2.16]	1.17 [3.37]
Controls	Yes	Yes	Yes
Style Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
Adj-R <sup>2</sup>	0.12	0.16	0.29
No. of Funds	5500	3146	1079
No. of Obs	23399	4676	1217
Mean(Scaled Residual)	-0.40	-0.68	-1.12

Table 1.7: Regression results with different cutoff thresholds.

early in its life. In comparison, funds with more latent risk would incur large negative scaled residuals much later due to the rare occurrence. This could produce a negative coefficient on age (but no effect on  $Age_{i,t}^2$ ). But none of these can explain the joint significance of 5 predictors across different subsamples. Thus, we believe our results are robust and consistent with what our model suggests.

## 1.9 Conclusion

In this paper, we analyze the impact of latent risk when asset managers are incentivized to manipulate performance in order to attract fund inflow. By solving a Bayesian rational equilibrium, in which investors correctly conjecture managers' manipulation strategy and manager correctly conjecture investors' evaluation rule, we find that both skilled and unskilled managers take on excessive latent risk to boost performance and engage in "noisy pooling" even if investors rationally update on crash loss. This leads to large welfare loss to both investors and managers and calls for third-party certification to investors or strictly enforced investment mandate. Also, wasteful excess risk taking reduces the marginal benefit of higher alpha, which discourages skilled manager from seeking costly alpha-generating opportunities and aggravates the slow-moving capital problem. It also significantly lengthens the time required by investors to discover a manager's ability. Fi-

Table 1.8: "Live" and "Graveyard" Subsamples		
	Dependent Variable: Scaled Residual	
	(Live)	(Graveyard)
Residual Vol	0.372	0.397
	[3.38]	[2.92]
Ability Difference	-1.32	-1.34
	[-5.54]	[-3.29]
Age	-0.230	-0.518
	[-3.09]	[-4.70]
Age <sup>2</sup>	0.010	0.020
	[2.30]	[3.39]
log(AUM)	0.024	0.818
	[0.41]	[2.57]
Controls	Yes	Yes
Style Fixed Effect	Yes	Yes
Year Fixed Effect	Yes	Yes
Adj-R <sup>2</sup>	0.16	0.21
No. of Funds	1568	1578
No. of Obs	2404	2272

Table 1.8: Regression results for (1.27) with "Live" and "Graveyard" subsamples.

nally, we show that the incentive for asset managers to take excessive latent risk could be an important factor that leads to the subprime mortgage crisis and provide a complementary explanation for declining return to scale in fund management industry.

There are several directions to extend our analysis. In our current model, the intensity of crash is fixed. It would be interesting to see the effect of endogenous choice of crash intensity as, in reality, managers do have freedom along this dimension. It might be possible that the skilled managers choose to have zero crash intensity to signal investors of their type. Also, our model abstract away from risk management. It is plausible to assume that a skilled manager may be good at both alpha-generation and risk management. In this case, she might be able to reduce the loss size upon crash. This could lead to higher latent risk exposure taken by the skilled managers than that taken by the unskilled at all reputation levels. Also, one would wish to know that if excessive latent risk taking can be handled with optimal contract design. Finally, asset pricing implications could be explored in a general equilibrium framework that endogenizes the latent risk premium.



## Appendix

### Proof of Proposition 2:

Conditioned on no jump

$$E_t^I \left[ \frac{dR_t}{R_t} | i = H \right] = E_t^I \left[ \left( \mu^H + \pi \hat{\phi}_t^H (\hat{p}_t) \right) dt + \sigma dB_t \right] = \left( \mu^H + \pi \hat{\phi}_t^H \right) dt$$

Similarly,

$$E_t^I \left[ \frac{dR_t}{R_t} | i = L \right] = E_t^I \left[ \left( \mu^L + \pi \hat{\phi}_t^L (\hat{p}_t) \right) dt + \sigma dB_t \right] = \left( \mu^L + \pi \hat{\phi}_t^L \right) dt$$

Hence,

$$\begin{aligned} \beta(\hat{p}_t) &= \frac{\hat{p}_t (1 - \hat{p}_t) \left[ \left( \mu^H + \pi \hat{\phi}_t^H \right) - \left( \mu^L + \pi \hat{\phi}_t^L \right) \right] dt}{\sigma^2 dt} \\ &= \frac{\hat{p}_t (1 - \hat{p}_t) \left[ \alpha + \pi \hat{\Delta} \right]}{\sigma^2} \end{aligned}$$

Also,

$$\begin{aligned} E_t^I [dR_t] &= \hat{p}_t E_t^I [dR_t | i = H] + (1 - \hat{p}_t) E_t^I [dR_t | i = L] \\ &= \bar{\mu} + \hat{p}_t \left( \alpha + \hat{\phi}^H (\hat{p}_t) \right) + (1 - \hat{p}_t) \hat{\phi}^L (\hat{p}_t) \end{aligned}$$

Thus,

$$\begin{aligned} d\hat{p}_t &= \beta(\hat{p}_t) (dR_t - E_t^I [dR_t]) \\ &= \beta(\hat{p}_t) \left( \alpha^i + \pi \hat{\phi}_t^i + \sigma dB_t - \left[ \hat{p}_t \left( \alpha + \pi \hat{\phi}^H \right) + (1 - \hat{p}_t) \left( \pi \hat{\phi}^L \right) \right] \right) \end{aligned}$$

### Proof of Proposition 3:

From investors' point of view, conditioned on  $i = H$ ,  $\varepsilon = \ln \left( 1 - \Lambda \hat{\phi}^H \right) - \ln \chi - L$ . Since  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ , this occur with probability  $\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left( -\frac{(\ln(1-\Lambda\hat{\phi}^H) - \ln\chi - L)^2}{2\sigma_\varepsilon^2} \right) d\varepsilon$ . Similarly, conditioned on  $i = L$ ,  $\varepsilon = \ln \left( 1 - \Lambda \hat{\phi}^L \right) - \ln \chi - L$ , which occur with probability

$$\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left( -\frac{(\ln(1-\Lambda\hat{\phi}^L) - \ln\chi - L)^2}{2\sigma_\varepsilon^2} \right) d\varepsilon.$$

Using Bayes rule, we have the after-jump

$$\begin{aligned}
\hat{p}_t &= P(\hat{p}_{t-}, \chi) \\
&= \frac{p_{t-} \exp \left( -\frac{(\ln(1-\Lambda\hat{\phi}^H) - \ln \chi - L)^2}{2\sigma_\varepsilon^2} \right) d\varepsilon}{p_{t-} \exp \left( -\frac{(\ln(1-\Lambda\hat{\phi}^H) - \ln \chi - L)^2}{2\sigma_\varepsilon^2} \right) d\varepsilon + (1-p_{t-}) \exp \left( -\frac{(\ln(1-\Lambda\hat{\phi}^L) - \ln \chi - L)^2}{2\sigma_\varepsilon^2} \right) d\varepsilon} \\
&= \frac{p_{t-}}{p_{t-} + (1-p_{t-}) \exp \left( \frac{1}{2\sigma_\varepsilon^2} \ln \left( \frac{1-\Lambda\hat{\phi}^H}{1-\Lambda\hat{\phi}^L} \right) \left[ \ln(1-\Lambda\hat{\phi}^H)(1-\Lambda\hat{\phi}^L) - 2\ln \chi - 2L \right] \right)}
\end{aligned}$$

**Proof of Proposition 4:**

Write the value function as  $V^i(W_t, \hat{p}_t, t) = \max_{\phi_{[t, \infty]}^i} E_t \left[ \int_t^\infty e^{-\rho s} \frac{W_s^{1-\gamma}}{1-\gamma} ds \right]$  subject to

$$\begin{aligned}
\frac{dW_t}{W_t} &= (\mu^i + \pi\phi_t^i + f(\hat{p}_t)) dt + \sigma dB_t + [(1-\Lambda\phi_t^i) e^{-L-\varepsilon} - 1] dN_t \\
d\hat{p}_t &= \beta(\hat{p}_t) \left( \alpha^i + \pi\phi_t^i + \sigma dB_t - \left[ \hat{p}_t (\alpha + \pi\hat{\phi}^H) + (1-\hat{p}_t) (\pi\hat{\phi}^L) \right] \right) \\
&\quad + (P(\hat{p}_{t-}, \phi_t^i, \varepsilon) - \hat{p}_{t-}) dN_t
\end{aligned}$$

Then the HJB equation is

$$\begin{aligned}
0 &= \max_{\phi_t^i} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} dt + V_t^i dt + V_W^i E_t[dW_t] + V_{\hat{p}}^i E_t[d\hat{p}_t] + V_{W\hat{p}}^i dW_t d\hat{p}_t + \frac{1}{2} V_{WW}^i (dW_t)^2 \\
&\quad + \frac{1}{2} V_{\hat{p}\hat{p}}^i (d\hat{p}_t)^2 + \lambda E_t(V^i((1-\Lambda\phi_t^i) e^{-L-\varepsilon} W_t, P(\hat{p}_{t-}, \phi_t^i, \varepsilon), t) - V^i) dt
\end{aligned}$$

Conjecture that the value function is the form  $V^i(W_t, \hat{p}_t, t) = \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U^i(\hat{p}_t)$  and substitute into the HJB equation above

$$\begin{aligned}
0 &= \max_{\phi_t^i} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} - e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} U^i + \frac{1}{\rho} e^{-\rho t} W_t^{1-\gamma} (\mu^i + \pi\phi_t^i + f(\hat{p}_t)) U^i + \\
&\quad \frac{1}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} \beta(\hat{p}_t) \left( \alpha^i + \pi\phi_t^i - \left[ \hat{p}_t (\alpha + \pi\hat{\phi}^H) + (1-\hat{p}_t) (\pi\hat{\phi}^L) \right] \right) U_{\hat{p}}^i \\
&\quad + \frac{1}{\rho} e^{-\rho t} \beta \sigma^2 W_t^{1-\gamma} U_{\hat{p}}^i - \frac{1}{2\rho} e^{-\rho t} \gamma \sigma^2 W_t^{1-\gamma} U^i + \frac{1}{2\rho} e^{-\rho t} \sigma^2 \beta^2 \frac{W_t^{1-\gamma}}{1-\gamma} U_{\hat{p}\hat{p}}^i + \\
&\quad \frac{\lambda}{\rho} e^{-\rho t} \frac{W_t^{1-\gamma}}{1-\gamma} \left[ e^{-(1-\gamma)L} (1-\Lambda\phi_t^i)^{1-\gamma} E[e^{-(1-\gamma)\varepsilon} U^i] - U^i(P) \right]
\end{aligned}$$

which simplifies to

$$\begin{aligned}
0 &= \min_{\phi_t^i} \rho - \left[ \rho + \lambda - (1 - \gamma) \left( \mu^i - \frac{1}{2} \gamma \sigma^2 + \pi \phi_t^i + f(\hat{p}_t) \right) \right] U^i + \\
&\quad \beta \left( \alpha^i + (1 - \gamma) \sigma^2 + \pi \phi_t^i - \left[ \hat{p}_t \left( \alpha + \pi \hat{\phi}^H \right) + (1 - \hat{p}_t) \left( \pi \hat{\phi}^L \right) \right] \right) U_{\hat{p}}^i + \\
&\quad \frac{1}{2} \sigma^2 \beta^2 U_{\hat{p}\hat{p}}^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^i)^{1-\gamma} E \left[ e^{-(1-\gamma)\varepsilon} U^i(P) \right]
\end{aligned}$$

Collecting terms related to  $\phi_t^i$ , we have

$$\phi_t^i = \arg \min \pi \left[ (1 - \gamma) U^i + \beta U_{\hat{p}}^i \right] \phi^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^i)^{1-\gamma} E \left[ e^{-(1-\gamma)\varepsilon} U^i(P) \right]$$

**Proof of Proposition 5:**

Using results from proposition 3, and since  $\beta = 0$  and  $P(\hat{p}_t, \phi_t^i, \varepsilon) = \hat{p}_t$  (0 or 1) the minimization over  $\phi_t^i$  becomes much easier.

$$\phi_t^i = \arg \min \pi \left[ (1 - \gamma) U^i \right] \phi^i + \lambda e^{-(1-\gamma)L + \frac{1}{2}(1-\gamma)^2 \sigma_\varepsilon^2} (1 - \Lambda \phi^i)^{1-\gamma} U^i$$

F.O.C gives

$$0 = \pi (1 - \gamma) - \lambda \Lambda (1 - \gamma) e^{-(1-\gamma)L + \frac{1}{2}(1-\gamma)^2 \sigma_\varepsilon^2} (1 - \Lambda \phi^i)^{-\gamma}$$

which yields

$$\phi^i = \frac{1 - e^{\frac{\gamma-1}{\gamma} L + \frac{(\gamma-1)^2}{2\gamma} \sigma_\varepsilon^2} \left( \frac{\lambda \Lambda}{\pi} \right)^{\frac{1}{\gamma}}}{\Lambda} = \bar{\phi} \quad (1.34)$$

Also, we have

$$\lambda \Lambda (1 - \gamma) e^{-(1-\gamma)L + \frac{1}{2}(1-\gamma)^2 \sigma_\varepsilon^2} (1 - \Lambda \bar{\phi})^{-\gamma} = \pi (1 - \gamma)$$

Since  $\beta = 0$ , we have

$$\begin{aligned}
0 &= \rho - \left[ \rho + \lambda - (1 - \gamma) \left( \mu^i - \frac{1}{2} \gamma \sigma^2 + \pi \bar{\phi} + f(0) \right) \right] U^i(0) \\
&\quad + \lambda e^{-(1-\gamma)L} (1 - \Lambda \bar{\phi})^{1-\gamma} E \left[ e^{-(1-\gamma)\varepsilon} U^i(0) \right]
\end{aligned}$$

and

$$\begin{aligned}
0 &= \rho - \left[ \rho + \lambda - (1 - \gamma) \left( \mu^i - \frac{1}{2} \gamma \sigma^2 + \pi \bar{\phi} + f(1) \right) \right] U^i(1) \\
&\quad + \lambda e^{-(1-\gamma)L} (1 - \Lambda \bar{\phi})^{1-\gamma} E \left[ e^{-(1-\gamma)\varepsilon} U^i(1) \right]
\end{aligned}$$

These imply that

$$\begin{aligned} U^i(0) &= \frac{\rho}{\rho + \lambda - (1 - \gamma) \left( \mu^i - \frac{1}{2} \gamma \sigma^2 + \pi \bar{\phi} + f(0) \right) - \lambda e^{-(1-\gamma)L} (1 - \Lambda \bar{\phi})^{1-\gamma} e^{\frac{1}{2}(1-\gamma)^2 \sigma^2}} \\ &= \frac{\rho}{\rho + \lambda + (\gamma - 1) \left( \mu^i + \pi \bar{\phi} + f(0) - \frac{1}{2} \gamma \sigma^2 \right) - \frac{\pi}{\Lambda} (1 - \Lambda \bar{\phi})} \end{aligned} \quad (1.35)$$

Similarly,

$$U^i(1) = \frac{\rho}{\rho + \lambda + (\gamma - 1) \left( \mu^i + \pi \bar{\phi} + f(1) - \frac{1}{2} \gamma \sigma^2 \right) - \frac{\pi}{\Lambda} (1 - \Lambda \bar{\phi})} \quad (1.36)$$

### Numerical Procedures For Solving The Fixed-Point:

We shall solve the fixed-point problem through discretization and iteration, which is similar to Tauchen and Hussey (1991). Since the state variable  $\hat{p}_t \in [0, 1]$ , we shall discretize state space into equally spaced vector  $[p_0, p_1, \dots, p_k, \dots, p_N] = [0, \delta, 2\delta, \dots, 1 - \delta, 1]$ . Then denote  $U^i(p_k)$  as  $U_k^i$  with  $k = 0, 1, \dots, N$ .

Since the differential equations (1.20) and (1.21) contain integral parts (expectations), we shall discretize and transform the integral into summation.

#### 1. Change of Measure

Before we can do that, we need to remove the  $e^{-(1-\gamma)\varepsilon}$  that appears inside the expectation through change of measure. More specifically,

$$\begin{aligned} & E_t \left[ e^{-(1-\gamma)\varepsilon} U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon)) \right] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}} e^{-(1-\gamma)\varepsilon} U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon)) d\varepsilon \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{\varepsilon^2 + 2(1-\gamma)\sigma_\varepsilon^2\varepsilon + (1-\gamma)^2\sigma_\varepsilon^4}{2\sigma_\varepsilon^2}} e^{\frac{1}{2}(1-\gamma)^2\sigma_\varepsilon^2} U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon)) d\varepsilon \\ &= e^{\frac{1}{2}(1-\gamma)^2\sigma_\varepsilon^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{(\varepsilon + (1-\gamma)\sigma_\varepsilon^2)^2}{2\sigma_\varepsilon^2}} e^{\frac{1}{2}(1-\gamma)^2\sigma_\varepsilon^2} U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon)) d\varepsilon \\ &= e^{\frac{1}{2}(1-\gamma)^2\sigma_\varepsilon^2} \tilde{E}_t [U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon))] \text{ where } \varepsilon \sim N((\gamma - 1)\sigma_\varepsilon^2, \sigma_\varepsilon^2) \end{aligned} \quad (1.37)$$

#### 2. Change of Variable

The state price  $\hat{p}_t$  is between 0 and 1. Notice that fixing  $\hat{p}_{t-}$  and  $\phi_{t-}^i$ ,  $P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon)$  maps  $\varepsilon \in R$  into  $[0, 1]$ . Thus, temporarily treating  $\hat{p}_{t-}$  and  $\phi_{t-}^i$  as constants, which are already fixed at the time of jump, we can let  $q := P(\varepsilon)$ . The density of  $q$  and density of  $\varepsilon$  are related through the change of variable formula:

$$f_q(q) = f_\varepsilon(P^{-1}(q)) \left| \frac{dP^{-1}(q)}{dq} \right|$$

where  $f_q$  and  $f_\varepsilon$  denotes the density functions for  $q$  and  $\varepsilon$ .

Since  $P(\varepsilon)$  is given by

$$\frac{p_{t-}}{p_{t-} + (1 - p_{t-}) \exp \left( \frac{1}{2\sigma_\varepsilon^2} \ln \left( \frac{1 - \Lambda \hat{\phi}_{t-}^H}{1 - \Lambda \hat{\phi}_{t-}^L} \right) \left[ \ln \frac{(1 - \Lambda \hat{\phi}_{t-}^H)(1 - \Lambda \hat{\phi}_{t-}^L)}{(1 - \Lambda \hat{\phi}_{t-}^i)^2} + 2\varepsilon \right] \right)}$$

we have

$$\begin{aligned} \varepsilon &= P^{-1}(q) \\ &= \left[ \ln \left( \frac{\hat{p}_{t-}}{1 - \hat{p}_{t-}} \right) + \ln \left( \frac{1 - q}{q} \right) \right] \frac{\sigma_\varepsilon^2}{\hat{G}^-} - \frac{\hat{G}^+}{2} + \ln(1 - \Lambda \hat{\phi}_{t-}^i) \quad (1.38) \end{aligned}$$

where  $\hat{G}^- = \ln(1 - \Lambda \hat{\phi}_{t-}^H) - \ln(1 - \Lambda \hat{\phi}_{t-}^L)$   
 $\hat{G}^+ = \ln(1 - \Lambda \hat{\phi}_{t-}^H) + \ln(1 - \Lambda \hat{\phi}_{t-}^L)$

$$\frac{dP^{-1}(q)}{dq} = -\frac{\sigma_\varepsilon^2}{\hat{G}^-} \frac{1}{q(1-q)}$$

This allows us to fully characterize  $f_q(q)$ .

### 3. Approximating $\tilde{E}[U^i(P(\hat{p}_{t-}, \phi_{t-}^i, \varepsilon))]$

We let  $q$  fall onto the same grid of  $\hat{p}_t$ . Then, we can write

$$\tilde{E}[U^i(P)] \approx \frac{\sum_k f_q(q_k) U_k^i}{\sum_k f_q(q_k)} \quad (1.39)$$

bearing in mind that  $f_q$  depends also on  $\hat{p}_{t-}$  and  $\phi_{t-}^i$ , which have been treated as constant at the moment of jump.

### 4. Discretize the differential-integral equation

Once we are able to express the integral part in a sum, discretizing the differential part is standard. Here let us take  $\phi^i(\cdot)$  to be a given function first. We approximate

$$\begin{aligned} U_{\hat{p}}^i(p_k) &\approx \frac{U_{k+1}^i - U_{k-1}^i}{2\delta} \\ U_{\hat{p}\hat{p}}^i(p_k) &\approx \frac{U_{k+1}^i - 2U_k^i + U_{k-1}^i}{\delta^2} \end{aligned}$$

The ODE becomes a system of linear equations: for each  $k$

$$-\rho = K_{0,k}^i U_k^i + K_{1,k}^i \frac{U_{k+1}^i - U_{k-1}^i}{2\delta} + K_{2,k}^i \frac{U_{k+1}^i - 2U_k^i + U_{k-1}^i}{\delta^2} + K_{3,k}^i \frac{\sum_l f_q(q_k; p_k, \phi^i(\cdot)) U_k^i}{\sum_l f_q(q_k; p_k, \phi^i(\cdot))} \quad (1.40)$$

where

$$\begin{aligned}
K_{0,k}^i &= -\rho - \lambda + (1 - \gamma) \left( \mu^i + \pi \phi^i(p_k) + f(p_k) - \frac{1}{2} \gamma \sigma^2 \right) \\
K_{1,k}^H &= \beta(p_k) [\alpha(1 - p_k) + (1 - \gamma) \sigma^2 + \pi(1 - p_k) \Delta(p_k)] \\
K_{1,k}^L &= \beta(p_k) [-\alpha p_k + (1 - \gamma) \sigma^2 - \pi p_k \Delta(p_k)] \\
K_{2,k}^i &= \frac{1}{2} \sigma^2 \beta(p_k)^2 \\
K_{3,k}^i &= \lambda e^{-(1-\gamma)L + \frac{1}{2}(1-\gamma)^2 \sigma_\varepsilon^2} (1 - \Lambda \phi^i(p_k))^{1-\gamma}
\end{aligned}$$

This allows us to solve for the  $(N - 1) \times (N - 1)$  linear system and find out  $U_1^i, \dots, U_{N-1}^i$  ( $U_0^i = U^i(0)$  and  $U_N^i = U^i(1)$  are known boundary conditions).

##### 5. Policy Iteration For Fixed-Point

In step 4, we take  $\phi^i(\cdot)$  as given. The value function computed  $U^i$  is the value function of type- $i$  manager in the situation where investors conjecture  $\phi^i(\cdot)$  and manager of type  $i$  follows  $\phi^i(\cdot)$  (not necessarily optimal of manager). To find the fixed-point in which investors conjecture  $\phi^i(\cdot)$  and manager of type  $i$  follows  $\phi^i(\cdot)$ , which happens to be optimal for him, we need to employ policy iteration with slight modification. Standard policy iteration for dynamic programming should fix investors conjecture  $\hat{\phi}^i(\cdot)$  first and iterate manager's policy through calculating  $U^i$  and computing the optimal policy based on  $U^i$  in an alternating manner until convergence. Then use the converged  $\phi^i(\cdot)$  as the next conjecture by investors. But this is too time-consuming. Here, we follow the steps below:

- (0) Start with a conjecture  $\phi_{(0)}^i(\cdot)$  and set iteration number  $n = 0$
- (1) For a given conjecture  $\phi_{(n)}^i(\cdot)$ , calculate the value function  $U_{(n)}^i$  if manager follows  $\phi_{(n)}^i(\cdot)$
- (2) Based investors' conjecture  $\phi_{(n)}^i(\cdot)$  and value function  $U_{(n)}^i$ , find the optimal policy  $\tilde{\phi}_{(n+1)}^i(p_k)$  by numerically minimizing

$$\pi [(1 - \gamma) U^i(p_k) + \beta U_p^i(p_k)] \phi^i + \lambda e^{-(1-\gamma)L} (1 - \Lambda \phi^i)^{1-\gamma} \tilde{E} [U^i(P(p_k, \phi^i, \varepsilon))]$$

- (3) If  $\|\phi_{(n)}^i - \tilde{\phi}_{(n+1)}^i\|$  is less than some specified error, we can stop. Otherwise, we use the weighted Jacobi procedure by choosing a  $w < 1$  and setting

$$\phi_{(n+1)}^i = w \phi_{(n)}^i + (1 - w) \tilde{\phi}_{(n+1)}^i$$

to ensure smooth convergence. With  $\phi_{(n+1)}^i$  calculated, go back to step (1).

## References

- Adelino, M. 2010, "Do investors rely only on ratings? The case of mortgage-backed securities," *Working Paper*
- Aggarwal, V., N. D. Daniel, N. Naik 2009, "Role of managerial incentives and discretion in hedge fund performance," *Journal of Finance*, 64, 2221-2256
- Arora, N., N.J. Ju, and H. Ou-Yang 2006, "A model of asset pricing under portfolio delegation and differential information," *Working Paper*
- Basak, S., A. Pavlova, and A. Shapiro 2007, "Optimal asset allocation and risk shifting in money management," *Review of Financial Studies*, 20, 1583-1621
- Basak, S., A. Pavlova, and A. Shapiro 2008, "Offsetting the implicit incentives: Benefits of benchmarking in money management," *Journal of Banking & Finance*, 32, 1883-1893
- Basak, S., and D. Makarov 2009, "Strategic asset allocation in money management," *Working Paper*
- Berk, J. and R. Green 2004, "Mutual fund flows and performance in rational markets." *Journal of Political Economy*, 112, 1269-1295
- Bollen, N. and R. Whaley 2009, "Hedge fund risk dynamics: Implications for performance appraisal," *Journal of Finance*, 64, 985-1035
- Brown, S., W. Goetzmann, B. Liang and C. Schwartz 2008, "Mandatory disclosure and operational risk: Evidence from hedge fund registration," *Journal of Finance*, 63, 2785-2815
- Brunnermeier, M., S. Nagel and L. Pedersen 2008, "Carry trades and currency crashes," *NBER Macroeconomics Annual*, 23, 313-347
- Brunnermeier, M. and L. Pedersen 2009, "Market liquidity and funding liquidity," *Review of Financial Studies*, 22, 2201-2238
- Cadenillas, A., J. Cvitanic and F. Zapareto 2007, "Optimal risk-sharing with effort and project choice," *Journal of Economic Theory*, 133, 403-440
- Cao, C., Y. Chen, B. Liang and A. W. Lo 2010, "Can hedge funds time market liquidity?" *Working Paper*
- Carpenter, J. 2000, "Does Option Compensation Increase Managerial Risk Appetite?" *Journal of Finance*, 55, 2311-2331
- Chapman, D. A., R. B. Evans and Z. Xu 2010, "The portfolio choices of young and old active mutual fund managers," *Working Paper*

- Chen, J., H. Hong, M. Huang, J. D. Kubik 2004, "Does fund size erode mutual fund performance? The role of liquidity and organization," *American Economic Review*, 94, 1276-1302
- Chevalier, J. and G. Ellison 1997, "Risk taking by mutual funds as a response to incentives," *Journal of Political Economy*, 105, 1167-1200
- Chevalier, J. and G. Ellison 1999, "Career concerns of mutual fund managers," *Quarterly Journal of Economics*, 114, 389-432
- Cole, H. L. and M. Obstfeld 1991, "Commodity trade and international risk sharing: How much do financial markets matter?" *Journal of Monetary Economics*, 28, 3-24
- Coval, J. D., J. Jurek, E. Stafford 2007, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 86, 479-512
- Coval, J. D., J. Jurek, E. Stafford 2009, "Economic catastrophe bonds," *American Economic Review*, 99, 628-666
- Cuoco, D. and R. Kaniel 2010, "Equilibrium prices in the presence of delegated portfolio management," Forthcoming *Journal of Financial Economics*
- Dasgupta, A. and A. Prat 2006, "Financial equilibrium with career concerns," *Theoretical Economics*, 1, 67-93
- Dasgupta, A. and A. Prat 2008, "Information aggregation in financial markets with career concerns," *Journal of Economic Theory*, 143, 83-113
- Dow, J. and G. Gorton 1997, "Noise trading, delegated portfolio management, and economic welfare", *Journal of Political Economy*, 105, 1024-1050
- Dybvig, P., H. Farnsworth and J. Carpenter 2010, "Portfolio performance and agency," *Review of Financial Studies*, 23, 1-23
- Fung, W. and D. Hsieh 2004, "Hedge fund benchmarks: a risk based approach," *Financial Analyst Journal*, 60, 65-80
- Getmansky, M. 2005, "The life cycle of hedge funds: Fund flows, size and performance," *Working Paper*
- Getmansky, M., A. W. Lo and I. Makarov 2004, "An econometric model of serial correlation and illiquidity in hedge fund returns," *Journal of Financial Economics*, 74, 529-609
- Goetzmann, W., J. Ingersoll, M. Spiegel and I. Welch 2007, "Portfolio performance manipulation and manipulation-proof performance measures," *Review of Financial Studies*, 20, 1503-1546
- Guasoni, P., G. Huberman and Z. Wang, "Performance maximization of actively managed funds," Forthcoming *Journal of Financial Economics*



- Guerrieri, V. and P. Kondor 2010, "Fund managers, career Concerns, and asset price volatility," *Working Paper*
- He, Z. and W. Xiong 2009, "Dynamic debt run," *Working Paper*
- He, Z. and W. Xiong 2011, "Delegated asset management and investment mandate," *Working Paper*
- Hu, X., J. Pan and J. Wang 2010, "Noise as information for illiquidity," *Working Paper*
- Jensen, M. and W. M. Meckling 1976, "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics*, 3, 305-360
- Kaniel, R. and P. Kondor 2010, "The delegated Lucas tree," *Working Paper*
- Khandani, A. and A. W. Lo 2009, "What happens to the Quants in August 2007," *Working Paper*
- Liptser, R. and A. N., Shiryaev 1978, "Statistics of random processes," *Springer*
- Lo, A. W. 2001, "Risk management for hedge funds: Introduction and overview," *Financial Analysts Journal*, 57, 16-33
- Makarov, I. and G. Plantin 2010, "Rewarding trading skills without inducing gambling," *Working Paper*
- Malliaris, S., and H. Yan 2010, "Reputation concerns and slow-moving capital," *Working Paper*
- Mitchell, M. and T. Pulvino 2002, "Characteristics of risk and return in risk arbitrage," *Journal of Finance*, 56, 2135-2175
- Mitchell, M. L. Pedersen and T. Pulvino 2007, "Slow moving capital," *American Economic Review*, 97, 215-220
- Ou-Yang, H. 2003, "Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem," *Review of Financial Studies*, 16, 173-208
- Panageas, S. and M. M. Westerfield 2009, "High-water marks: high risk appetites? Convex compensation, long horizons, and portfolio choice," *Journal of Finance*, 64, 1-36
- Ross, S. A. 2004, "Compensation, incentives, and the duality of risk aversion and riskiness," *Journal of Finance*, 54, 207-225
- Sirri, E.R. and P. Tufano 1998, "Costly search and mutual fund flows," *Journal of Finance*, 53, 1589-1622
- Stein, J. and D. Scharfstein 1990, "Herd behavior and investment," *American Economic Review*, 80, 465-479

Tauchen, G. and R. Hussey 1991, "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models," *Econometrica*, 59, 371-396

Vayanos, D. and P. Woolley 2010, "An institutional theory of momentum and reversal," *Working Paper*

He, Z. and A. Krishnamurthy 2010, "Intermediary asset pricing," *Working Paper*

## Chapter 2

# Price Manipulation In General Equilibrium

(This is joint work with Jiang Wang)

### 2.1 Introduction

With the rise of delegated asset management over the past several decades, the presence of large institutional investors with significant price impact is well-recognized. While price impact has been largely considered as a curse in the studies of optimal execution strategies, the associated power to move prices has also captured researchers' attention. Indeed, quite a number of price manipulation schemes have been proposed featuring a large trader with no genuine trading needs, who simply tries to extract a positive expected profit by mimicking an informed trader, exploiting behavioral biases or sabotaging market-clearing/trading rules<sup>1</sup> etc. According to Allen and Gale (1992), they are broadly classified into "action-based", "information-based" and "trade-based". However, in these manipulation schemes, price-setting power is always coupled with additional market imperfections such as elaborate market structure or information asymmetry. One naturally wonders could price-setting power alone lead to successful manipulation?

To address this question, we study price manipulation in a general equilibrium framework, in which the only market imperfection is the presence of a non-competitive large trader. Competitive traders are fully rational and there is no information asymmetry about future payoffs or the identity of the large trader. Market-clearing and trading rules are also standard: the large trader determines the security prices and clears the market. Surprisingly, we find that, even in a static framework, there could be price manipulation that is "malicious" in nature.

A large trader may have genuine trading need that allows for potential Pareto improvement. They manipulate security prices only to seek better transaction prices and take a greater share of the risk-sharing benefit. This type of manipulation is "benign" as argued

---

<sup>1</sup>For instance, market cornering and short squeeze.

in Kyle and Viswanathan (2008). To distinguish malicious manipulation from benign ones, we start with an intuitive notion of "no trading motive". There is no trading motive if a large trader would not trade were she to behave competitively. If the large trader possesses no trading motive but still wishes to trade (necessarily improving her welfare), she is considered to be engaging in "pure manipulation". "Pure manipulation" is malicious in the sense that, with no trading motive, there is already no room for greater allocative efficiency through risk-sharing. While a large trader behaving competitively cannot improve her welfare through trading, a non-competitive large trader manages to do so. The only source of the large trader's welfare improvement is her manipulation power.

We begin our investigation by proving the existence of a "large trader equilibrium" in a static framework under no further assumptions that are needed to ensure the existence of a competitive equilibrium. We then find that the possibility of pure manipulation is equivalent to the failure of the Weak Axiom of Revealed Preference (WARP) for the aggregate security demand of competitive traders at the competitive equilibrium without the large trader. Such a failure essentially means that, under some other security price, the competitive traders in aggregate would hold a portfolio worth negative value priced using the competitive equilibrium prices. If WARP at competitive equilibrium price holds, there is no pure manipulation whereas, if it fails, a sufficiently risk-tolerant large trader can conduct pure manipulation. By definition, pure manipulation requires no trading motive, which implies that the large trader's marginal utility valuation of the security payoffs (without any trading) is the same as the prices in the competitive equilibrium. If WARP at competitive equilibrium fails, the large trader can manipulate security prices to a new one and squeeze out a change in portfolio holdings by the competitive traders that has negative value and, therefore, leads to an utility improvement for large trader. A sufficient condition for the failure of WARP at competitive equilibrium prices is that the local security demand is jointly upward-sloping in some portfolio of securities. Most interestingly, this also guarantees the possibility of pure manipulation of arbitrarily small scale. In reality, even a large institutional investor has capital constraint and, therefore, limited power to move market prices. Also, under SEC regulation, one has to report within 10 days any ownership of an equity security greater than 5%. Price manipulation achieved through large trade would be quickly under the scrutiny of the regulators. Pure manipulation of arbitrarily small scale circumvents these issues and is, therefore, hard to eradicate.

Our analysis links price manipulation to a rich literature in general equilibrium theory. Based on the existing results and some new findings of our own, we identify three important factors that leads to pure manipulation. Firstly, we need heterogeneity in risk preferences and endowments of the competitive traders. Secondly, we need large trading needs in the competitive equilibrium among competitive traders. Finally, we need large remaining insuring needs even after competitive risk-sharing. These three factors contribute to a positive aggregate wealth effect that dominates the substitution effect, which always keeps the demand for any portfolio of securities jointly downward-sloping. Under the broad set of expected-utility preferences, we find that there are abundant cases that lead to jointly upward-sloping security demand.

With CRRA utilities, "peso problem" that involves a disaster state in aggregate endowments could lead to pure manipulation. Heterogeneous endowment allocations lead to heterogeneous wealth effect of consumption demand with respect to changes in relative prices. Heterogeneous risk preferences and large remaining insurance needs ensure that the aggregate wealth effect is positive. Large trading needs guarantee that substitution effect of aggregate consumption demand is dominated by the aggregate wealth effect. This gives rise to a jointly upward-sloping consumption demand and potential pure manipulation opportunities.

Finally, we extend our setup to a multi-period economy. The transformation of a multi-period problem into a static one is well-known in competitive equilibrium studies. Indeed, if the large trader is able to commit to a price manipulation plan in the beginning, the multi-period economy is no different from a static one. However, in general, the large trader is unable to commit if there is dynamic trading over time, which is a typical problem faced by a intertemporal monopolist. Rather than choosing manipulation plan freely as in a multi-period commitment equilibrium, the large trader is constrained by the time-consistency requirement placed by the multi-period dynamic equilibrium. Thus, "going dynamic" only reduces the manipulation power as the large trader effectively chooses from a subset of price manipulation plans. This suggests that we could always use manipulation in the static framework (equivalent to commit equilibrium in multi-period economy) as an upper bound for the severity of price manipulation. Moreover, an implication is that while it is commonly thought that different security structures that complete the markets give rise to the same equilibrium state prices and allocations, this no longer holds when a large trader is introduced with no commitment ability in a multi-period economy with dynamic trading. For instance, a multi-period economy with contingent claims traded only in the beginning is equivalent to a commitment equilibrium, which yields greater manipulation power for the large trader than other security structures that involves tradings at different times<sup>2</sup>.

This paper proceeds as follows. Section 1.1 reviews related literature. Section 2 studies the large trader equilibrium in a static economy. Section 2.1 discusses the model setup. Section 2.2 studies the existence and welfare properties of a general large trader equilibrium. Section 2.3 focuses on pure manipulation and elaborates on conditions that give rise to it in complete markets. Section 3 extends our analysis to a multi-period economy. Section 3.1 introduces further notations for a multi-period setup. Section 3.2 analyzes the impact of different security structures with the same asset span. Section 3.3 shows that results in a static economy carry over to a multi-period large trader commitment equilibrium with some minor modifications. Section 4 concludes.

### 2.1.1 Related Literature

Price manipulation by a non-competitive investor has been studied before. On the "benign" side of price manipulation, Kyle (1985) provides a classic example of how a large

---

<sup>2</sup>For e.g. dynamic trading of a long-lived stock and a short-term bond.

trader with genuine informational trading motive splits her trades over time to reduce price impact. Vayanos (2001) studies how a large trader with risk-sharing motive trades against competitive traders. With information asymmetry about endowment shocks, his study also touches on the potential "round-trip" trades that aims at misleading competitive traders. Basak (1997) analyzes the behavior of a large trader sharing risk with competitive traders in a complete markets general equilibrium framework and focuses on how the endowment risk of the large trader gets incorporated into the pricing of a risky asset. These studies involve a large trader with genuine trading motive and, therefore, do not fall into the class of pure manipulation, which our paper focuses on.

On the more "malicious" side of price manipulation, many mechanisms have been proposed. Allen and Gale (1992) categorizes these into "information-based", "action-based" and "trade-based". "Informed-based" manipulation requires potential information advantage and is implemented by misleading investors through false signal (Vila (1989), Benabou and Laroque (1992), Van Bommel (2003)) or trade-disclosure (Fishman and Hagerty (1995), John and Narayanan (1997)). "Action-based" encompasses actions that change the actual or perceived value of the underlying firms. Bagnoli and Lipman (1996) considers a fake take-over bid. Goldstein and Guebel (2008) studies a case, in which the large trader, through short-selling, misleads the firm managers to forgo positive NPV investment opportunities. Among "trade-based" manipulation schemes, a classical example is Allen and Gale (1992), which demonstrates how an uninformed large trader, by mimicking an informed one, could extract a positive expected profit from competitive traders, who cannot tell the identity of the large trader. Allen and Gorton (1992) finds, with more liquidity-driven trades among sell orders than buy orders, asymmetric impact on prices leads to manipulation opportunities. Mei, Wu and Zhou (2004) shows that a large trader can exploit the disposition effects among competitive traders to generate positive manipulation profit. Kumar and Seppi (1992) formulates a manipulation scheme targeted at liquidity difference between derivatives and spot markets. Jarrow (1992, 1994) analyzes mechanisms such as non-synchronization across security markets and market cornering with subsequent short squeezes that engender manipulation opportunities. Khwaja and Mian (2005), Aggarwal and Wu (2006), Allen, Litov and Mei (2008) and Eom, Lee and Park (2009) provide several interesting empirical studies on price manipulations in financial markets of developed and developing countries. We find that additional market imperfections such as information asymmetry, behavioral biases or malfunctioning market-clearing mechanisms are not necessary. Pure manipulation could happen under standard trading and market-clearing procedure with fully rational and symmetrically informed competitive traders. Instead, a new mechanism involving heterogeneity in preferences and endowments, large trading needs and significant remaining insurance demand is proposed. Kyle and Viswanathan (2008) defines "illegal manipulation" that hinders both informational and allocative efficiency. Pure manipulation is less restrictive in comparison.

Our paper is related to the studies in the security structures (Hart (1975), Kreps (1982), Duffie and Huang (1985)). While these studies focus on competitive equilibria, we introduce a large trader with price-setting ability. In particular, we show that security structures that allow dynamic trading opportunities over time reduce the manipulation power of a

large trader due to the commitment problem. In addition, our paper is related to the general equilibrium theories. A branch of this literature has studied non-competitive producers, who engage in Bertrand or Cournot competition to maximize profit (see Hart (1985) and Bonanno (1994) for a survey). We prove the existence of a large trader equilibrium and analyze the welfare implications under general security structures. With incomplete markets, each security is a bundle of consumption (possibly negative) at different states and financial markets differ from goods markets fundamentally. Stahn (1998) and Giraud and Stahn (2003) study the generic existence of non-competitive equilibria with incomplete markets. However, their work requires more elaborate set of assumptions whereas ours require no assumptions beyond those that guarantee the existence of a competitive equilibrium. Finally, our work touches on the security demand curve (Shleifer (1986)) and heterogeneous risk-aversion (Wang (1996), Chan and Kogan (2002)). We show that, with sufficient heterogeneity in risk preferences and endowments, the security demand curve might not be downward-sloping.

## 2.2 Static Economy

In this section, we shall investigate into a 2-period pure-exchange economy with a single consumption good. Trading takes place at time-0 and assets pay off at time-1. As will be discussed in greater details in the next section, results in this section will largely carry into a multi-period economy when the large trader is able to commit to her strategy at time-0, which happens to deliver the greatest manipulation power.

### 2.2.1 Model Setup

There are  $K + 1$  states of nature.  $\omega_0$  is the time-0 state and  $\{\omega_1, \omega_2, \dots, \omega_K\}$  are  $K$  time-1 states, uncertainty of which will be revealed at time-1. We assume that  $\omega_k$  will occur with probability  $p_k = \mathbb{P}(\omega_k)$ . As a result, we must have  $p_0 = 1$  and  $\sum_{k=1}^K p_k = 1$ .

#### *Securities*

We assume that there are  $M$  proper securities available for trading at time-0 indexed by  $\{1, \dots, M\}$ . Security  $m$  pays  $X_{km}$  in state  $k$  for  $k = 0, 1, \dots, K$ . Hence, its payoff vector is  $X_m = (0 \ X_{1m} \ X_{2m} \ \dots \ X_{Km})^\top$ .<sup>3</sup> For convenience, we shall treat 1 unit of time-0 consumption itself as a security indexed by 0. Its payoff vector  $X_0 = (1 \ 0 \ \dots \ 0)^\top$ . By making such a treatment, prices of securities are no longer in terms of time-0 consumption. They become completely nominal and can be scaled up or down by any positive constant without affecting the portfolio choices. Putting all the securities together, we have the augmented asset payoff matrix of dimension  $(K + 1) \times (M + 1)$ :

$$X^+ = (X_0 \ X_1 \ X_2 \ \dots \ X_M)$$

---

<sup>3</sup>The first entry is 0 for the  $M$  securities since they do not pay anything at time-0. It is slightly unconventional to include time-0 payoff. But it turns out to be more convenient in our analysis.

The nominal time-0 prices of securities are determined endogenously (up to a positive scaling constant) in equilibrium and denoted as  $S = (S_0 \ S_1 \dots S_M)^\top$  where  $S_m$  is the price of security  $m$ . Given that we treat time-0 consumption as a security on its own,  $S_0$  might not be 1. However, we can always scale the price vector  $S$  so that  $S_0 = 1$  without creating any changes to the equilibrium quantities. By doing so, we come back to the conventional setup. Also, note that, at this stage, we do not assume complete market. As a result, there is no restriction on the number of available proper securities relative to the number of states in time-1.

#### *Agents' Preferences and Endowments*

We assume there are two types of agents in the economy. The first type are  $N$  competitive trader, who are price-taking. The second type is a large trader, denoted as  $L$ , who is non-competitive. For  $i = 1, 2, \dots, N, L$ , trader  $i$  has von Neumann-Morgenstern utility function  $U^i$  over time-0 and time-1 consumptions. Given trader  $i$ 's consumption  $c^i = (c_0^i \ c_1^i \dots c_K^i)^\top$ , her expected utility is

$$U^i(c^i) = u^i(c_0^i) + \sum_{k=1}^K u^i(c_k^i) p_k$$

with  $u^i(\cdot)$  being the Bernoulli utility function. For simplicity, we ignore the intertemporal discount factor. Trader  $i$  is assumed to have endowment  $e^i = (e_0^i \ e_1^i \dots e_K^i)^\top$ . We do not require  $e^i$  to lie in the span of asset payoff matrix.

There are three conventional assumptions imposed on the utility function and endowments of competitive traders  $i = 1, \dots, N$  to ensure the existence of a competitive equilibrium without the presence of the large trader<sup>4</sup>:

- (1)  $u^i(c) : \mathbb{R}_+ \rightarrow \mathbb{R}$  is continuous and infinitely differentiable on  $\mathbb{R}_{++}$ ,  $u_c^i > 0$  and  $u_{cc}^i < 0$
- (2) if  $c \in \mathbb{R}_{++}^{K+1}$ ,  $\{c' \in \mathbb{R}^{K+1} : U^i(c') \geq U^i(c)\} \subseteq \mathbb{R}_{++}^{K+1}$
- (3)  $e^i \in \mathbb{R}_{++}^{K+1}$

Assumption (1) specifies the usual properties of utility function: agents have to consume non-negative amount at each state and their preferences are strictly monotonically increasing and concave. Assumption (2) prevents the solution of the agent's maximizing problem from occurring at the boundary. Assumption (3) asserts that each competitive trader's endowment at each state is positive. Furthermore, let us assume that  $u^L(c)$  is smooth and increasing. Also,  $U^L(e^L) > -\infty$ , which suggests that the large trader is able to "survive" without having to trade with competitive traders. We shall denote the absolute and relative risk-aversion of trader  $i$  as  $\alpha^i(c) = -u_{cc}^i(c)/u_c^i(c)$  and  $\gamma^i(c) = -u_{cc}^i(c)c/u_c^i(c)$  respectively.

#### *Trading Mechanism and Market Clearance*

We assume that large trader has the freedom to set the price of securities but is under the obligation to fulfill aggregate security demand by competitive traders. Specifically, the trading mechanism at time-0 is as follows: the large trader announces to security

<sup>4</sup>See Magill and Quinzii (1996) and Debreu (1972).



prices  $S$  to competitive traders. Competitive trader  $i$  submits her security demand  $\theta^i(S) \in \mathbb{R}^{M+1}$  with  $\theta_m^i(S)$  being her demand for security  $m$ . The aggregate security demand by all competitive traders is  $\theta(S) = \sum_{i=1}^N \theta^i(S)$ . The large trader has to take the opposite position  $\theta^L = -\theta(S)$  to clear the market. Price-setting ability could be achieved by submitting market order<sup>5</sup>. Indeed, "price-setting" and "submitting market order" can be viewed as the two sides of the same coin if the market demand curve by competitive traders is monotonically downward/upward-sloping. However, if there are multiple market-clearing security prices that correspond to a market order by the large trader, the "submitting market order" approach becomes problematic. For simplicity, we shall assume that the large trader sets the securities prices and act as the trading counterparty to fulfill competitive traders' security demand. This is closer to the reality with a large market-maker quoting prices to small investors and absorbing the aggregate order flow. With a few exceptions such as small-cap stocks, complete power to set price at wish is impossible. However, it is not difficult to imagine that certain institutional players with enormous amount of capital such as prominent investment banks or large pension funds do enjoy the freedom to locally move the price to their desire. We shall discuss local price manipulation later on as well.

#### *Equilibrium*

We shall define the set of security prices that will lead to arbitrage opportunities for competitive traders as

$$\mathcal{A} = \{S \in \mathbb{R}^{M+1} : \exists \theta \in \mathbb{R}^{M+1} \text{ s.t. } X^+ \theta - (S^\top \theta \ 0 \ \dots \ 0)^\top > 0\}$$

Clearly, the large trader will never pick a security price vector  $S$  that belongs to the set  $\mathcal{A}$ . By doing so, she will have to fulfill arbitrage positions taken by competitive traders and incur infinite amount of loss. This leads us to conclude that any redundant security can be removed from the set of securities since its price would have to be the price of the replicating portfolio so as to avoid arbitrage. Therefore, we can assume, with loss of generality, that there is no redundant security among the  $M$  proper securities:

$$\dim \langle X_1, \dots, X_M \rangle = M$$

Competitive traders take security price vector  $S$  as given. As a result, they solve a utility maximization problem:

$$\max_{\theta^i} U^i(e^i + X^+ \theta^i) \text{ s.t. } S^\top \theta^i \leq 0 \quad (2.1)$$

With endowment  $e^i$  and security demand  $\theta^i$ , competitive trader  $i$ 's consumption will be  $c^i = e^i + X^+ \theta^i$ . But she must satisfy her budget constraint: the nominal value of her total securities demand (including the time-0 consumption) must be non-positive. Assuming that such an optimal portfolio  $\theta^i(S)$  exists (the existence shall be discussed in the next subsection) in response to a given security price  $S$ , the total security demand by competitive traders is  $\theta(S) = \sum_{i=1}^N \theta^i(S)$ , which the large trader has to fulfill by trading

---

<sup>5</sup>See e.g. Basak (1997), Vayanos (2001) etc.

a portfolio  $\theta^L(S) = -\theta(S)$ . The resulting consumption by the large trader is

$$c^L = e^L + X^+\theta^L(S) = e^L - X^+\theta(S)$$

Taking the aggregate security demand function  $\theta(\cdot)$  as given, the large trader solves

$$\max_{S \in \mathbb{R}^{M+1} \setminus \mathcal{A}} U^L(e^L - X^+\theta(S)) \quad (2.2)$$

The large trader always stays away from the security prices that generate arbitrage. With competitive traders' aggregate security demand function in mind, she optimizes over security prices to obtain the aggregate security demand that delivers the highest utility for herself. Thus, a "large trader equilibrium" is defined as follows.

**Definition 1** A "large trader equilibrium" is defined as the equilibrium price  $\hat{S}$  such that

- (1) Competitive trader  $i$  submits optimal demand  $\theta^i(\hat{S})$  to maximize utility
- (2) The large trader trades  $\theta^L(\hat{S}) = -\sum_{i=1}^N \theta^i(S)$  to fulfill aggregate security demand by competitive traders.
- (3)  $\hat{S}$  maximizes large trader's utility

The existence of the "large trader equilibrium" will be shown in the next subsection. Besides, we shall introduce two hypothetical equilibrium notions that serve as comparisons to the "large trader equilibrium". The first is a competitive equilibrium with  $N$  competitive traders alone in the absence of the large trader, which we shall refer to as "competitive equilibrium without  $L$ ". The second is also a competitive equilibrium with  $N$  competitive traders as well as the large trader, who were to behave competitively. We shall refer to this equilibrium as the "competitive equilibrium with price-taking  $L$ ".

We shall introduce a few notations. Let  $c = \sum_{i=1}^N c^i$  and  $e = \sum_{i=1}^N e^i$  be the aggregate consumption and endowments of competitive traders. The excess consumption demand by competitive trader  $i$  is defined as  $z^i = c^i - e^i$ , which is the difference between her consumption and endowment. The aggregate excess consumption demand is defined as  $z = \sum_{i=1}^N z^i = c - e$ . Additionally, we shall write the indirect utility  $V^i(\theta^i)$  for trader  $i = 1, \dots, N, L$  as a function of her security demand  $\theta^i$ . Hence,  $V^i(\theta^i) = U^i(e^i + X^+\theta^i)$ . We shall denote equilibrium quantities in the "large trader equilibrium" as  $\hat{\cdot}$  in general<sup>6</sup>. Similarly, equilibrium quantities in the "competitive equilibrium without  $L$ " is denoted as  $\bar{\cdot}$ . Lastly, for any multivariate differentiable function  $f(x) : \mathbb{R}^r \rightarrow \mathbb{R}^s$ , we shall denote its  $s \times r$  matrix derivative as  $\partial f$ .

---

<sup>6</sup>For instance, equilibrium security price vector in the "large trader equilibrium" is denoted as  $\hat{S}$ .

### 2.2.2 General Manipulation

In this subsection, we shall study on the existence of a "large trader equilibrium" and some general properties under the "large trader equilibrium". "Pure manipulation", which is the most interesting special case of "large trader equilibrium", shall be analyzed in the next subsection.

When large trader chooses the optimal security prices to set, she needs to take into full consideration of the aggregate security demand  $\theta$  from the competitive traders, which she has to fulfill by taking on the opposite position  $\theta^L = -\theta$ . Thus, the existence of a well-behaved aggregate security demand  $\theta(S)$  as function of security prices  $S$  needs to be explored before we can embark on demonstrating the existence of the "large trader equilibrium". Fortunately, earlier studies have extensively analyzed the aggregate security demand. The key results are stated in the following proposition.

**Proposition 1** *Given the three assumptions about competitive traders' utility functions, endowments as well as the assumption about the non-redundancy of traded securities, the optimal security demand  $\theta^i$  by each competitive trader  $i$  exists and is unique if the security prices  $S$  does not present any arbitrage opportunities (i.e.  $S \notin \mathcal{A}$ ). This ensures that  $\theta^i(S)$  is well-defined on  $\mathbb{R}^{M+1} \setminus \mathcal{A}$ . Furthermore,  $\theta^i(S)$  is smooth on  $\mathbb{R}^{M+1} \setminus \mathcal{A}$ .*

**Proof.** See Magill and Quinzii (1996) for a proof. ■

The existence of a well-defined and smooth security demand function for each competitive trader on the set of no-arbitrage prices will lead to a well-defined and smooth aggregate security demand function for competitive traders  $\theta(S)$  (as well as  $\theta^L(S) = -\theta(S)$ ) on the set of no-arbitrage prices. With a well-behaved aggregate security demand function, large trader's optimization problem is well-defined. And the existence of "large trader equilibrium" hinges on the existence of a optimal security price vector that attains the highest utility for the large trader and is bounded in each coordinate.

**Proposition 2** *Under previous assumptions, large trader equilibrium exists.*

The intuition for its existence is straightforward. Large trader can easily avoid trading with competitive traders by setting the security prices to be the equilibrium prices in the "competitive equilibrium without  $L$ ", which yields zero aggregate security demand. As we assume that the large trader can survive without having to trade with competitive traders, this provides a feasible security price vector to set. By manipulating security prices, the large trader can at most take away all of the endowments of the competitive traders. This gives an upper bound on the attainable level of utility for the large trader. As a result, the highest attainable level of utility should exist and can be achieved by some security price vector. This particular security price vector will be the large trader equilibrium price. Of course, this price will not incur any arbitrage, which provides "free lunch" to competitive traders and results in infinite loss for the large trader.

The interesting question is how the large trader equilibrium is different from the competitive equilibrium with price-taking  $L$ . In a similar setup, Basak (1997) proves, under complete markets, that (1) the equilibrium quantities will be different as long as there is trading between the large trader and the competitive traders (2) whenever the equilibrium quantities in the large trader equilibrium is different, the large trader equilibrium is inefficient. We shall conduct similar analysis under general security structure (possibly incomplete) that will yield (1) and (2) in Basak (1997). Surprisingly, we shall demonstrate that the presence of the large trader will never be so detrimental to competitive traders that it yields a Pareto disimprovement from the competitive equilibrium without  $L$ .

We shall employ the notion "constrained efficiency"<sup>7</sup> when market is incomplete. More specifically, we have the following definition:

**Definition 2** *Efficiency (if market is incomplete, then constrained-efficiency): An allocation  $(c^1, \dots, c^N, c^L)$  with  $c^i \in \mathbb{R}_+^{K+1}$  is efficient if*

1.  $c^i = e^i + X^+ \theta^i$  for some  $\theta^i$  (Feasibility under the security structure)
2.  $\sum_{i=1, \dots, N, L} c^i \leq \sum_{i=1, \dots, N, L} e^i$  (Feasibility under endowments constraint)
3. There doesn't exist any allocation  $(x^1, \dots, x^N, x^L)$  such that  $U^i(x^i) \geq U^i(c^i) \forall i$  with strict inequality for at least one  $i$

It has been well-known that any competitive equilibrium is efficient (or constrained efficient) in a 2-period economy<sup>8</sup>. The efficiency of a large trader equilibrium is investigated in the following propositions.

**Proposition 3** *If there is trading between large trader and competitive traders (i.e.  $\hat{\theta}^L \neq 0$ ) in the large trader equilibrium, the allocation is not efficient. Otherwise (i.e.  $\hat{\theta}^L = 0$ ), the allocation is efficient.*

The above proposition suggests that the involvement of a large trader always leads to inefficiency unless her presence is muted by her own choice (i.e.  $\theta^L = 0$ ). The next natural question is when is a large trader's presence is not muted. Proposition 4 presents a simple sufficient condition.

**Proposition 4** *If there is trading between the large trader and competitive traders in the competitive equilibrium with price-taking  $L$ , there is always trading between the large trader and competitive traders in the large trader equilibrium.*

---

<sup>7</sup>See Diamond (1967).

<sup>8</sup>Competitive equilibrium may not be efficient for a multi-period economy if market is incomplete.

Proposition 3 and 4 allow us to characterize the efficiency implications of the large trader equilibrium in the Table 2.1. If the large trader trades in the competitive equilibrium with price-taking  $L$ , she will trade in the large trader equilibrium, which results in an inefficient allocation. If the large trader does not trade in the competitive equilibrium with price-taking  $L$ , she may or may not trade in the large trader equilibrium. If she trades, the allocation is inefficient; if she does not, the allocation is efficient. (1) in Basak (1997) can be seen from the fact that the large trader will trade in the large trader equilibrium if she trades in the competitive equilibrium with price-taking  $L$ . Since the allocation is inefficient when she trades in large trader equilibrium, the equilibrium allocation will certainly be different from the efficient one if she were to behave competitively. (2) in Basak (1997) follows directly from the fact that when the large trader trades in the large trader equilibrium, the allocation is inefficient.

Table 2.1: Efficiency Implications

		Competitive Equilibrium with Price-Taking $L$	
Large Trader Equilibrium		$L$ Trades	No Trade
	$L$ Trades	Inefficient	Inefficient
	No Trade	Impossible	Efficient

While the comparison with the competitive equilibrium with price-taking  $L$  serves as an indication of the detrimental effects of large trader manipulation, a comparison with the competitive equilibrium without  $L$  shows that the large trader's presence will not be so detrimental that the equilibrium allocation is a Pareto disimprovement for competitive traders, which somewhat bounds the severity of the manipulation problem. This is stated in Proposition 5.

**Proposition 5** *Not all competitive traders will have lower welfare in the large trader equilibrium than in the competitive equilibrium without  $L$ .*

To conclude this section, we state a trivial observations in the following corollary without proof.

**Corollary 1** *Whenever there is trade by the large trader in the large trader equilibrium, at least one of the competitive traders has lower welfare and the large trader has higher welfare relative to the competitive equilibrium with price-taking  $L$ .*

### 2.2.3 Pure Manipulation

As shown in Table 2.1, it is possible that the large trader (behaving competitively) trades in the competitive equilibrium with price-taking  $L$ . In this case, the underlying reason for her trade is genuine risk-sharing with the competitive traders. As Proposition 4 suggests, when she behaves non-competitively, she will still trade although the equilibrium allocation under her manipulation will be different. By behaving non-competitively, she engages in price manipulation. However, this form of manipulation is "benign" in the sense that it is undertaken to share risk and the presence of the large trader might benefit all competitive trader relatively to the competitive equilibrium without  $L$ . Manipulation simply serves to obtain better transaction prices and procure a greater share of the overall economic benefit generated. Such manipulation is similar to the form of large trader manipulation studied in Kyle (1985) or Vayanos (2001) with genuine informational or risk-sharing trading motives behind.

However, there exists the possibility that if the large trader were to behave competitively, she would not trade with the competitive traders. In this case, the prices in the competitive equilibrium with price-taking  $L$  would be the same as in the competitive equilibrium without  $L$ . By behaving non-competitively, the large trader may be able to move the prices away from those in the competitive equilibrium in order to induce trading for her own welfare improvement. We shall refer to this type of manipulation as "pure manipulation", which is not motivated by genuine risk-sharing trading motive.

First of all, we shall define what we mean by "no trading motive" formally.

**Definition 3** *Suppose that the prices in a competitive equilibrium without  $L$  is  $\bar{S}$ . If the large trader's endowment is such that  $\partial U^L(e^L) X^+ = \lambda \bar{S}^\top$  for some  $\lambda > 0$ , the large trader has **no trading motive** with respect to the competitive equilibrium without  $L$ . Otherwise, there is **trading motive**.*

If the large trader were to behave competitively,  $\partial U^L(e^L) X^+ = \lambda \bar{S}^\top$  would guarantee that large trader will submit zero security demand under the prevailing equilibrium prices  $\bar{S}$ . The equilibrium prices in the competitive equilibrium without  $L$ ,  $\bar{S}$ , would prevail because the security structure does not allow further risk-sharing between the large trader and competitive traders. If market is complete, the definition of "no trading motive" implies that  $\partial U^L(e^L) = \lambda^i \partial U^i(\bar{c}^i)$ , which means that the marginal utilities of all traders including the large one are completely aligned and proportional to each other. If  $\partial U^L(e^L) X^+ \neq \lambda \bar{S}^\top$ , the marginal utility valuation of the securities by the large trader is not the same as the competitive traders. This allows Pareto improvement through risk-sharing under the existing security structure, which is a genuine trading motive. With genuine trading motive, the large trader can always improve her utility through local manipulation, which we shall define below.

**Definition 4** *There exists local manipulation with respect to  $\bar{S}$  if, for arbitrarily small  $\varepsilon > 0$ , there exists a price vector  $S \in \mathbb{R}^{M+1}$  such that  $\|S - \bar{S}\|_2 < \varepsilon$  and the large trader improves her utility by setting security price vector to  $S$ .*

In reality, even a large institutional investor has capital constraint and, therefore, limited power to move market prices. Also, under SEC regulation, one has to report within 10 days any ownership of an equity security greater than 5%. Price manipulation of large scale would be quickly under the scrutiny of the regulators. Local manipulation allows arbitrarily small scale. It circumvents these issues and is, therefore, more relevant practically and harder to detect.

**Proposition 6** *If the large trader has trading motive with respect to  $\bar{S}$  (i.e.  $\partial U^L(e^L) X^+ \neq \lambda \bar{S}^\top$ ), there is local manipulation.*

The proposition above justifies our definition of "trading motive". With that, we can define pure manipulation as follows.

**Definition 5** *If the large trader has no trading motive with respect to  $\bar{S}$  but can manipulate the price to improve her utility, there is **pure manipulation**.*

If there are multiple competitive equilibria (with different equilibrium prices) without  $L$ , then there would always be pure manipulation. Suppose there are two equilibrium prices  $\bar{S}$  and  $\bar{S}'$  with  $\bar{S} \neq \bar{S}'$ . Since the large trader can only have no trading motive with respect to at most one competitive equilibrium, say  $\bar{S}$ , she always has trading motive with respect to  $\bar{S}'$ . Thus, if the prevailing equilibrium security price vector is  $\bar{S}$ , setting price to  $\bar{S}'$  will always improve her utility. But the existence of multiple competitive equilibria does not guarantee the possibility of local pure manipulation because, generically, equilibrium prices are finite.

It is important to note that pure manipulation can never lead to a Pareto improvement for the competitive traders relative to the competitive equilibrium without  $L$ . The reason is clear: if the large trader were to behave competitively, then she would not trade, which would leave the equilibrium consumptions of competitive traders unchanged. This is the competitive equilibrium with price-taking  $L$ . Since the competitive equilibrium with price-taking  $L$  is efficient, we cannot find a Pareto improvement. Pure manipulation involves trading between the large trader and competitive traders, which suggests that it is different from the competitive equilibrium with price-taking  $L$ . Therefore, this is certainly not a Pareto improvement for the competitive traders. Since the large trader gains from pure manipulation, some competitive trader must have lower utility under pure manipulation than in the competitive equilibrium with price-taking  $L$ , which implies that the same competitive trader has lower utility than in the competitive equilibrium without  $L$ . This contrasts with the case with trading motive, which allows for the possibility of

Pareto improvement for all competitive traders relative to the competitive equilibrium without  $L$ .

This observation leads us to conclude that pure manipulation is a form of "malicious" manipulation. In this sense, it is similar to the action-based, information-based or trade-based manipulation schemes studied in previous literatures, which also tend to extract benefit through trading with no genuine trading motive. However, it is important to note that pure manipulation does not require positive expected return, which is the objective of previous manipulation schemes. It is much milder in that it only requires the possibility to improve the utility of the large trader. With pure manipulation defined, we shall study under what conditions it is possible for the large trader to conduct it. It turns out that the possibility of pure manipulation is equivalent to the failure of the weak axiom of revealed preference (WARP) of the aggregate security demand at the competitive equilibrium without  $L$ . WARP is an important property of aggregate demand, which has been well-studied in the field of general equilibrium theory.

**Definition 6** *Aggregate security demand function  $\theta(S)$  of the competitive traders is said to satisfy the Weak Axiom of Revealed Preference if, for any pair of security price vectors  $S$  and  $S'$  such that  $\theta(S) \neq \theta(S')$ ,*

$$S^\top \theta(S') \leq 0 \text{ implies } S'^\top \theta(S) > 0$$

The underlying motivation is best understood at the level of individual trader.  $S^\top \theta^i(S') \leq 0$  suggests that  $\theta^i(S')$  is a feasible portfolio choice when the security price vector is  $S$ . However, since the optimal security demand submitted by trader  $i$  is  $\theta^i(S) \neq \theta^i(S')$ , this implies that portfolio  $\theta^i(S)$  is more desirable than  $\theta^i(S')$ . When security price vector is  $S'$ , the fact that trader  $i$  chooses  $\theta^i(S')$  rather than  $\theta^i(S)$  implies that  $\theta^i(S)$  must not be feasible under the new security price vector  $S'$  (i.e.  $S'^\top \theta(S) > 0$ ). WARP holds for every rational individual with consistent portfolio choice decisions and it certainly holds true in our case. However, it does not necessarily hold at the aggregate level since heterogeneity in preferences and endowments prevent the the aggregate security demand from behaving as if it were rational decisions by a single individual.

**Proposition 7** *If Weak Axiom of Revealed Preference (WARP) for aggregate security demand  $\theta(S)$  of competitive traders holds at the security price vector,  $\bar{S}$ , of competitive equilibrium without  $L$ , pure manipulation is not possible. Conversely, if it fails at  $\bar{S}$  with strict inequality (i.e. there exists a price vector  $S$  such that  $\bar{S}^\top \theta(S) < 0$ ), there exists a sufficiently risk-tolerant<sup>9</sup> large trader, who can conduct price manipulation to improve her utility.*

---

<sup>9</sup> By "sufficiently risk-tolerant", we mean  $\|\partial^2 U^L\|$  (taken under the matrix norm) is sufficiently small uniformly.



Proposition 7 links pure manipulation to the aggregate property of the security demand. By definition, pure manipulation requires no trading motive, which implies the marginal utility valuation by the large trader (without any trading) of the security payoffs is the same as (up to a positive scaling constant) the competitive equilibrium security price vector  $\bar{S}$ . If WARP fails at  $\bar{S}$ , it essentially means that the large trader can find another price  $S$  such that the aggregate change in portfolio holdings by competitive traders is worth a negative value priced under  $\bar{S}$ . The aggregate holdings by the large trader, which is the opposite to the aggregate change competitive traders' security demand by market-clearing, has a positive value under  $\bar{S}$ . Since  $\bar{S}$  is the same as marginal utility valuation by the large trader, the trade induced by moving security prices from  $\bar{S}$  to  $S$  will improve her utility as long as her marginal utility does not change too much when the associated consumption change takes place. This will be true for a sufficiently risk-tolerant large trader.

Note that we do not need WARP to hold everywhere to avoid pure manipulation. Rather, we only need it to hold at  $\bar{S}$ , the security prices of the competitive equilibrium without  $L$ . Also, if WARP fails at  $\bar{S}$ , the large trader can set a price to improve her utility provided she is sufficiently risk-tolerant. Proposition 7 allows us to connect our analysis to a rich body of literature in general equilibrium theory, which we shall discuss in more details later under complete markets. However, the failure of WARP at  $\bar{S}$  does not guarantee local pure manipulation opportunities. We shall specify a sufficient condition for local manipulation to exist in the proposition below.

**Proposition 8** *If  $\partial\theta(\bar{S})$  is NOT negative semi-definite, WARP at  $\bar{S}$  fails locally with strict inequality, by which we mean*

$$\forall \varepsilon > 0, \exists S \in \mathbb{R}^{M+1} \text{ such that } \|S - \bar{S}\|_2 < \varepsilon \text{ and } \bar{S}^\top \theta(S) < 0$$

*This allows a sufficiently risk-tolerant large trader to conduct local pure manipulation of arbitrarily small scale.*

Since  $\partial\theta(\bar{S})\bar{S} = -\theta(\bar{S}) = 0^{10}$ ,  $\partial\theta(\bar{S})$  is never negative definite. But it could be negative semi-definite. Proposition 8 suggests when it is not negative semi-definite, the large trader can set a price arbitrarily close to  $\bar{S}$  and improve her utility through local pure manipulation. The interpretation of non-negative semi-definite  $\partial\theta(\bar{S})$  is straightforward: non-negative semi-definiteness implies that there exists a direction of price change  $\Delta \in \mathbb{R}^{M+1}$  such that  $\Delta^\top \partial\theta(\bar{S}) \Delta > 0$ .  $\partial\theta(\bar{S}) \Delta$  is the first order effect of aggregate security demand change in response to the price change. Define a portfolio of securities  $\Psi = \partial\theta(\bar{S}) \Delta$ . The change in portfolio value caused by security prices change is  $\Delta^\top \Psi > 0$ . Hence, after we change the security prices so that the portfolio  $\Psi$ 's price is higher, the aggregate demand for this portfolio is also higher. This indicates a locally upward-sloping demand for portfolio  $\Psi$ . A simple illustration is given by the following corollary for a single-security portfolio.

---

<sup>10</sup>See Mas-Collel (1985) for a simple proof.

**Corollary 2** *If*

$$\max_{m \in \{0,1,\dots,M\}} \frac{\partial \theta_m(\bar{S})}{\partial S_m} > 0$$

, *there can be local pure manipulation.*

**Proof.** *If a diagonal entry is positive, the matrix is not negative semi-definite. ■*

This is equivalent to the demand curve of a particular security  $m$  being locally upward-sloping at  $\bar{S}$ . However, it is worth emphasizing that we do not need upward-sloping demand for any single security. Rather, the demand can be downward-sloping for every single security but still upward-sloping for a certain portfolio. The corollary below gives a two-securities portfolio case.

**Corollary 3** *If*

$$\min_{m_1, m_2 \in \{0,1,\dots,M\}} 4 \frac{\partial \theta_{m_1}(\bar{S})}{\partial S_{m_1}} \frac{\partial \theta_{m_2}(\bar{S})}{\partial S_{m_2}} - \left( \frac{\partial \theta_{m_1}(\bar{S})}{\partial S_{m_2}} + \frac{\partial \theta_{m_2}(\bar{S})}{\partial S_{m_1}} \right)^2 < 0$$

, *there can be local pure manipulation.*

**Proof.**  $\partial \theta + \partial \theta^\top$  is symmetric. And for a symmetric matrix to be negative semi-definite, any submatrix (with the omitted row and column indices being the same) must be negative semi-definite. If the condition in this corollary holds, there exists a  $2 \times 2$  submatrix

$$\begin{bmatrix} 2 \frac{\partial \theta_{m_1}}{\partial S_{m_1}} & \frac{\partial \theta_{m_1}}{\partial S_{m_2}} + \frac{\partial \theta_{m_2}}{\partial S_{m_1}} \\ \frac{\partial \theta_{m_1}}{\partial S_{m_2}} + \frac{\partial \theta_{m_2}}{\partial S_{m_1}} & 2 \frac{\partial \theta_{m_2}}{\partial S_{m_2}} \end{bmatrix}$$

with eigenvalues of opposite signs. So it is not negative semi-definite, which implies  $\partial \theta(\bar{S})$  is not negative semi-definite. ■

### A. Pure Manipulation In Complete Markets

Here we shall investigate pure manipulation in complete markets. By restricting ourselves to complete markets, the analysis becomes more tractable. Furthermore, in complete markets, consumption in each state of nature can be treated as a separate good. This allows us to make use of many results developed in the field of general equilibrium theory, which shed light on the potential reasons that give rise to the failure of WARP at competitive equilibrium without  $L$ .

It is well-known that, with complete markets, the exact security structure no longer matters for competitive equilibrium since all security structures can be reduced to contingent claims. We shall demonstrate that this is still true for the large trader equilibrium in the static economy. However, we shall show later that this is no longer the case for a multi-period dynamic economy.

**Proposition 9** *Given two different security structure  $X^+$  and  $X^{+'}$  that are of rank  $K + 1$  (i.e. complete markets), the large trader equilibrium (equilibria if non-unique) consumption allocations are the same.*

The underlying reason is straightforward. With complete markets, any non-arbitrage security price  $S \in \mathbb{R}^{M+1} \setminus \mathcal{A}$  will imply a particular state price vector  $\phi \in \mathbb{R}_{++}^{K+1}$  together with security structure  $X^+$ . Effectively, competitive traders' aggregate excess consumption is determined by the state price vector  $\phi$ . Hence, the large trader maximizes her utility by picking the best state price vector  $\hat{\phi}$ , which can then be implemented by a corresponding security price vector  $\hat{S} = \hat{\phi}^\top X^+$ . Thus, security structure does not matter and we can, without loss of generality, assume that the traded securities are contingent claims and the large trader manipulates state price vector  $\phi$  directly. As a result, the security demand and excess consumption are the same:  $\theta^i(\phi) = z^i(\phi)$ . The transformation of the problem establishes a connection to some well-known results that guarantee WARP at  $\bar{\phi}$ , the state prices in the competitive equilibrium without  $L$ .

**Proposition 10** *WARP holds at  $\bar{\phi}$ , the price in the competitive equilibrium without  $L$ , if*

1. *There exists a representative consumer<sup>11</sup>, which can be satisfied if*

- *Agents are homogeneous in preferences and endowments*
- *All competitive agents have preferences of the HARA form*

$$u^i(c) = \frac{(c - a^i)^{1-\gamma}}{1-\gamma} \text{ with the same } \gamma \text{ and possibly different } a^i \text{'s}$$

or

$$u^i(c) = -\exp(-\alpha^i c) \text{ with possibly different } \alpha^i \text{'s}$$

2. *Aggregate excess consumption  $z(\phi)$  exhibits **monotonicity condition**:*

$$[z(\phi) - z(\phi')]^\top (\phi - \phi') \leq 0 \text{ with strict inequality if } z(\phi) \neq z(\phi')$$

*which can be satisfied if  $\max_{c \in \mathbb{R}_{++}^{K+1}} \gamma^i(c) - \min_{c \in \mathbb{R}_{++}^{K+1}} \gamma^i(c) < 4 \forall i$  and endowments are collinear<sup>12</sup>.*

3. *Aggregate excess demand  $z$  exhibits **gross substitution**:*

$$\text{If } \phi_l > \phi'_l \text{ and } \phi_k = \phi'_k \forall k \neq l, z_k(\phi) > z_k(\phi') \forall k \neq l$$

*which can be satisfied if  $\gamma^i(c) < 1 \forall c \in \mathbb{R}_{++}^{K+1}$ .*

---

<sup>11</sup> Representative consumer's consumption choices coincide with the actual aggregate consumption choices at all prices.

<sup>12</sup> By "collinear", we mean the endowments of any two competitive traders  $i$  and  $j$  are proportional:  $e^i = k e^j$  for some constant scalar  $k > 0$ .

4. *There is no trading among competitive traders in the competitive equilibrium without  $L$ .*
5. *Complete insurance is reached without the large trader (regardless of preferences & endowment heterogeneity):*

*For all competitive traders,  $\bar{c}_k^i = \bar{c}_l^i$  for any two states  $k$  and  $l$*

*A necessary and sufficient condition for complete insurance is there is no aggregate risk:  $e_k = e_l$  for any two states  $k$  and  $l$ .*

Proposition 10 lists several sufficient conditions for WARP to hold at  $\bar{\phi}$ , which prevents pure manipulation. The breach of these conditions are necessary for pure manipulation opportunities to arise. Proposition 10.1 suggests that if competitive traders are sufficiently similar in terms of preferences and/or endowments so that there exists a representative consumer that rationalizes the aggregate consumption choices, there cannot be pure manipulation. Note that the requirement for a representative consumer to exist is much stronger than the requirement for a representative agent, which can be constructed for any competitive equilibrium. The key difference is that the representative agent's consumption choices coincide with the actual aggregate consumption choices (arising from summation of individual consumption choices) only at the equilibrium prices whereas the representative consumer's consumption choices coincide with the actual aggregate one at all prices. Proposition 10.1 indicates that we will not have pure manipulation with CRRA (Constant Relative Risk Aversion) utility function of the same relative risk-aversion or CARA (Constant Absolute Risk Aversion) utility functions.

Proposition 10.2 suggests we will not have pure manipulation if competitive traders' relative risk aversion does not vary too much and, more importantly, their endowments are proportional. The typical asset-pricing setup in finance literature with CRRA agents' and endowments as shares of a dividend stream falls into this category. Under CRRA utilities,  $\gamma^i(c)$  is a constant. Thus,  $\max_{c \in \mathbb{R}_{++}^{K+1}} \gamma^i(c) - \min_{c \in \mathbb{R}_{++}^{K+1}} \gamma^i(c) = 0 < 4$  is automatically satisfied. If agents' endowments are shares of a dividend stream, their endowments are collinear. These two pieces together prevent pure manipulation. The monotonicity condition essentially requires the uncompensated demand curve to be downward-sloping.

Proposition 10.3 introduces another condition that guarantees no pure manipulation. Gross substitution requires that a rise in one state price  $\phi_l$  (holding state prices at the other states fixed) will lead to higher demand for consumption at other states. By homogeneity, we can actually show that gross substitution also implies that such a price change will lead to a drop in consumption in state  $l$ :  $z_l(\phi) < z_l(\phi')$ <sup>13</sup>. In some sense, this is another version of downward-sloping demand (with substitution among consumptions at different states in addition). It guarantees WARP at  $\bar{\phi}$  rather than globally. The sufficient condition for

---

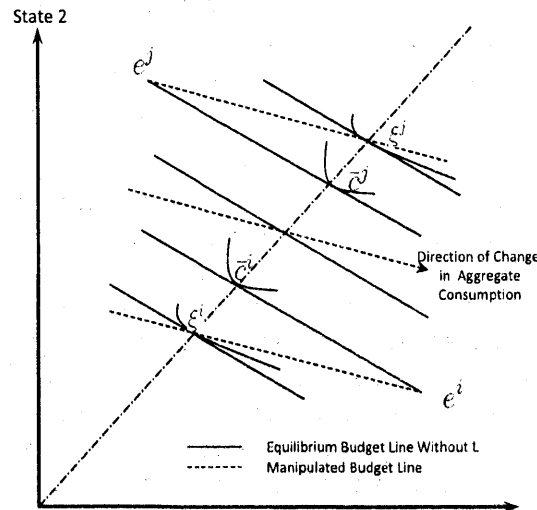
<sup>13</sup>See Mas-Collel, Whinston and Green (1995) for a proof.

gross substitution,  $\gamma^i(c) < 1$ , restricts competitive traders' relative risk aversion to be between 0 and 1, which again points towards sufficiently similar risk preferences.

Proposition 10.4 specifies that if the original endowments of the competitive traders are Pareto-optimal, WARP holds at  $\bar{\phi}$  and there is no pure manipulation. This is intuitive: since a competitive trader can always consume her endowment at any given prices, the large trader cannot lower any competitive trader's utility through price manipulation. But, as we have shown in Corollary 1, whenever the larger trader trades, at least one of the competitive agent will have lower utility. So the large trader will not trade and there will not be any price manipulation.

The intuition for Proposition 10.5 is illustrated by Fig 2.1 below with two competitive traders ( $i$  and  $j$ ) and two states of nature. With any budget line under manipulated prices, competitive agent  $i$  could trade to  $\xi^i$ , a riskless consumption plan. Since the gradient of indifference curves under expected utility are all the same at riskless consumption plans (45 degree line) regardless of utility functions or consumption levels, the gradient of indifference curve of agent  $i$  will be the same as the gradient of indifference curve at  $\bar{c}^i$  (also on 45 degree line), which has the same gradient as the budget line in the competitive equilibrium without  $L$ . Because the budget line under manipulated prices has a different gradient, it is not tangent to the indifference curve going through  $\xi^i$ . This implies that  $\xi^i$  is not a optimal choice. So both of them will trade to a more favorable consumption plan than  $\xi^i$ , which requires higher wealth evaluated under  $\bar{\phi}$ . But  $\sum \xi^i = \sum e^i$ . So if large trader manipulates  $\phi$  so that the budget line changes, competitive traders' aggregate consumptions will be worth more than their endowments evaluated at  $\bar{\phi}$ . Therefore, WARP holds at  $\bar{\phi}$  and there is no pure manipulation.

Figure 2-1: Complete Insurance



Proposition 10.1-3 all hinge on the requirement of similar competitive trader (either in preferences or in endowments) for WARP to hold at  $\bar{\phi}$ . Proposition 10.4-5 offer a different

angle: trading needs and remaining insurance demand. Proposition 10.4 suggests that pure manipulation has to require trading among competitive traders without the presence of the large trader. The security markets have to be vital for competitive traders. Proposition 10.5 puts forth the role of incomplete insurance in the origin of pure manipulation. If there is no aggregate risk, which allows riskless consumption plans to be attained by competitive traders without the large trader, pure manipulation is impossible. Remaining demand for insurance across states in the competitive equilibrium is necessary. Putting together, we see that heterogeneity among competitive traders, trading needs as well as remaining insurance demand in the competitive equilibrium without  $L$  are necessary conditions for pure manipulation to exist. Their presence allows the large trader to disrupt the allocative role of the security markets through price manipulation and gain benefit without genuine trading motive.

Proposition 8 specifies a sufficient condition under general security structure for local pure manipulation to exist. Under complete markets, this condition is equivalent to  $\partial z(\bar{\phi})$  being not negative semi-definite. Complete markets allow us to re-write a competitive trader  $i$ 's optimization problem as

$$\max_{c^i} U(c^i) \text{ s.t. } \phi^\top c^i \leq w^i \quad (2.3)$$

where  $w^i = \phi^\top e^i$  is the wealth of trader  $i$ . Thus,  $c^i(\phi) = \zeta^i(\phi, \phi^\top e^i)$  where  $\zeta^i(\phi, w^i)$  is trader  $i$ 's consumption under state prices  $\phi$  with wealth  $w^i = \phi^\top e^i$ . Thus,

$$\begin{aligned} \partial z^i &= \partial_\phi \zeta^i(\phi, \phi^\top e^i) + \partial_w \zeta^i(\phi, \phi^\top e^i) e^{i\top} \\ &= \underbrace{\partial_\phi \zeta^i(\phi, \phi^\top e^i) + \partial_w \zeta^i(\phi, \phi^\top e^i) e^{i\top}}_{\text{Substitution Effect}} \quad \underbrace{-\partial_w \zeta^i(\phi, \phi^\top e^i) z^{i\top}}_{\text{Wealth Effect}} \end{aligned} \quad (2.4)$$

Substitution matrix measures the effect on excess consumption with respect to price changes when wealth is compensated to make the consumption plan before price change still on the budget line. It is well-known from demand theory that the substitution matrix is always negative semi-definite<sup>14</sup>. As  $\partial z = \sum_{i=1}^N \partial z^i$ , it is clear from this analysis that what prevents  $\partial z$  from being negative semi-definite is the second part of (2.4) in aggregate  $\sum_{i=1}^N -\partial_w \zeta^i(\phi, \phi^\top e^i) z^{i\top}$ , which is a measure of wealth effect induced by price changes. If there is no trading needs,  $z^i(\bar{\phi}) = 0$  and the wealth effect is completely muted. If complete insurance is reached across states, it can be easily shown that  $\partial_w \zeta^i(\bar{\phi}, \bar{\phi}^\top e^i) = (1 \ p_1 \ \dots \ p_K)^\top$  for all  $i$ . Since, in equilibrium,  $z(\bar{\phi}) = \sum z^i(\bar{\phi}) = 0$ , the wealth effect is fully cancelled out among competitive traders. The conditions in Proposition 10.1-3 is about homogeneity among competitive traders, which again leads to  $\partial_w \zeta^i(\phi, \phi^\top e^i)$  to be close to each other and cancellation of wealth effect among competitive traders to a large extent. We shall further explore the above decomposition later for competitive traders with CRRA utilities in greater details.

Local pure manipulation of arbitrarily small scale is possible when  $\partial z(\bar{\phi})$  is not negative

<sup>14</sup>See MasColllel, Whinston and Green (1995).

semi-definite. Complete markets allow us to find explicit expression for the derivative of the excess consumption with respect to state price vector. Since  $\partial z(\phi) = \sum_{i=1}^N \partial z^i(\phi)$ , we shall find the derivative of excess consumption of each individual competitive trader first.

**Proposition 11** *For competitive trader  $i$ , the derivative of her state- $k$  consumption with respect to state- $j$  price is:*

$$\frac{\partial z_k^i}{\partial \phi_j} = \frac{\frac{1}{\alpha^i(c_j^i)} - z_j^i}{\alpha^i(c_k^i) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha^i(c_l^i)} \right)} \text{ if } j \neq k \quad (2.5)$$

$$\begin{aligned} \frac{\partial z_k^i}{\partial \phi_k} &= \frac{\frac{1}{\alpha^i(c_k^i)} - z_k^i}{\alpha^i(c_k^i) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha^i(c_l^i)} \right)} - \frac{1}{\alpha^i(c_k^i)} \frac{1}{\phi_k} \text{ if } j = k \\ &= -\frac{\frac{1}{\phi_k} \sum_{l \neq k} \frac{\phi_l}{\alpha^i(c_l^i)} + z_k^i}{\alpha^i(c_k^i) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha^i(c_l^i)} \right)} \end{aligned} \quad (2.6)$$

Note Proposition 11 holds at all state price vectors  $\phi \in \mathbb{R}_{++}^{K+1}$ . We do not need the equilibrium market-clearing condition  $\sum_{i=1}^N z^i = 0$  to reach it. Also, interestingly, (2.5) and (2.6) do not involve probability directly. Hence, heterogeneous beliefs do not have any effect on  $\partial z$  once consumption and state prices are fixed. It is easy to see that there is little restriction on the derivative of individual prices since  $\alpha^i(c) = -u_{cc}^i(c)/u_c^i(c)$  can be chosen arbitrarily. The first order condition of individual optimization imposes that

$$\partial U^i(c^i) = (u_c^i(c_0^i) \ u_c^i(c_1^i) \ \dots u_c^i(c_K^i)) = \lambda^i \pi^\top$$

where  $\pi$  is the vector of state price density. This only places restriction on the value of the first derivative of  $u^i$  at her chosen consumptions. However, other than being negative, there is no restriction on the second derivative of  $u^i$  at any given level of consumption. This gives us tremendous freedom to define the absolute risk aversion function  $\alpha^i(c)$ . It can be any smooth function that takes positive values. Therefore, local pure manipulation arise easily. We shall study local pure manipulation with the following two examples: one with general preferences and one with more specific CRRA utilities. The former allows us to see the abundance of local pure manipulation opportunities with the freedom to arbitrarily define absolute risk aversion function for competitive traders. Of course, one would question the existence of local pure manipulation opportunities with more restrictive set of utility function and the exact behavior of the aggregate excess consumption of competitive traders. These shall be addressed by the second example using CRRA utilities.

The tractability of CRRA utilities also helps us gain insights into causes of non-negative semi-definite  $\partial z(\bar{\phi})$  that leads to pure manipulation.

### General Preferences

We know from Corollary 2 that local pure manipulation exists as long as a diagonal entry of  $\partial z$  is positive, which corresponds to a upward-sloping demand for a single security. We shall consider a simple case with 2 competitive traders ( $i = 1, 2$ ) and show that demand could be locally upward-sloping. The derivative of the aggregate excess consumption for diagonal entries evaluated at  $\bar{\phi}$  is

$$\begin{aligned} \frac{\partial z_j(\bar{\phi})}{\partial \phi_j} &= \sum_{i=1,2} \frac{\partial z_j^i(\bar{\phi})}{\partial \phi_j} \\ &= - \frac{\left[ \left( \frac{1}{\phi_j} \sum_{l \neq j} \frac{\bar{\phi}_l}{\alpha^1(\bar{c}_l^1)} + \bar{z}_j^1 \right) \alpha^2(\bar{c}_j^2) \left( \sum_{l=0}^K \frac{\bar{\phi}_l}{\alpha^2(\bar{c}_l^2)} \right) + \right. \\ &\quad \left. \left( \frac{1}{\phi_j} \sum_{l \neq j} \frac{\bar{\phi}_l}{\alpha^2(\bar{c}_l^2)} + \bar{z}_j^2 \right) \alpha^1(\bar{c}_j^1) \left( \sum_{l=0}^K \frac{\bar{\phi}_l}{\alpha^1(\bar{c}_l^1)} \right) \right]}{\alpha^1(\bar{c}_j^1) \left( \sum_{l=0}^K \frac{\bar{\phi}_l}{\alpha^1(\bar{c}_l^1)} \right) \alpha^2(\bar{c}_j^2) \left( \sum_{l=0}^K \frac{\bar{\phi}_l}{\alpha^2(\bar{c}_l^2)} \right)} \end{aligned} \quad (2.7)$$

The denominator is always positive. Thus, it is the numerator that determines the sign of  $\frac{\partial z_j}{\partial \phi_j}$ . In particular, for  $\frac{\partial z_j}{\partial \phi_j} > 0$ , we want the numerator to be less than 0. The scaling property of  $z$  suggests that, without loss of generality, we can let the equilibrium price  $\bar{\phi}_j = 1$ . Also, note that  $\bar{z}_j^1 + \bar{z}_j^2 = 0$ . Upon simplifying, the numerator becomes

$$\begin{aligned} &\frac{(\rho_j^2(\bar{c}^2) \cdot \bar{\phi}) (\rho_j^1(\bar{c}^1) \cdot \bar{\phi} - 1)}{\alpha^1(\bar{c}_j^1)} + \frac{(\rho_j^1(\bar{c}^1) \cdot \bar{\phi}) (\rho_j^2(\bar{c}^2) \cdot \bar{\phi} - 1)}{\alpha^2(\bar{c}_j^2)} \\ &+ z_j^1 (\rho_j^2(\bar{c}^2) - \rho_j^1(\bar{c}^1)) \cdot \bar{\phi} \end{aligned} \quad (2.8)$$

where  $\rho_j^i(c^i) = \left( \frac{\alpha^i(c_j^i)}{\alpha^i(c_0^i)} \dots \frac{\alpha^i(c_K^i)}{\alpha^i(c_0^i)} \right)^\top$  is the vector of ratios of risk aversion at state- $j$  to risk aversion at state- $k$  for  $k = 0, 1, \dots, K$ . Since  $\rho_j^1(\bar{c}^1) \cdot \bar{\phi} - 1 = \sum_{l \neq j} \frac{\alpha^1(\bar{c}_l^1) \bar{\phi}_l}{\alpha^1(\bar{c}_j^1)} > 0$ , the first two terms are positive. The third term could be negative, which requires  $|z_j^1|$  and  $\|\rho_j^2(\bar{c}^2) - \rho_j^1(\bar{c}^1)\|_2$  to be large. In addition,  $z_j^1$  and  $(\rho_j^2(\bar{c}^2) - \rho_j^1(\bar{c}^1)) \cdot \bar{\phi}$  need to have opposite signs. In the two competitive traders case,  $|z_j^1| = |z_j^2|$  is the trade amount. This echos our previous analysis that large trading needs are needed for upward-sloping demand. Large  $\|\rho_j^2(\bar{c}^2) - \rho_j^1(\bar{c}^1)\|_2$  requires sufficiently different risk attitudes for trader 1 and trader 2 as well as sufficiently different consumption levels across the states for each of them<sup>15</sup>. These again reflect the need for preference heterogeneity and remaining insurance demand across states.

For a simple numerical illustration, assume that there is no actual consumption at time-0

<sup>15</sup>In the extreme case that the levels of consumption are the same across the states,  $\rho_j^i(\bar{c}^i) = (1 \ 1 \ \dots 1)^\top$ , which leads to perfect cancellation.

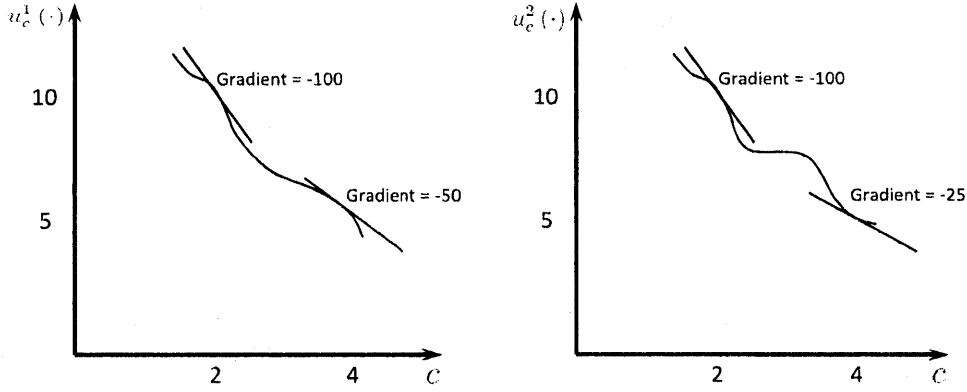


and there are two states at time-1 ( $k = 1, 2$ ) with equal probabilities  $p_1 = p_2 = \frac{1}{2}$ . We can let  $j = 1$  and further reduce (2.8) to

$$\bar{\phi}_2 \left[ \left( \frac{[\rho_1^1(\bar{c}^1)]_2 \bar{\phi}_2 + 1}{\alpha^2(\bar{c}_2^2)} + \frac{[\rho_1^2(\bar{c}^2)]_2 \bar{\phi}_2 + 1}{\alpha^1(\bar{c}_2^1)} \right) + \bar{z}_1^1 ([\rho_1^2(\bar{c}^2)]_2 - [\rho_1^1(\bar{c}^1)]_2) \right]$$

where  $[\rho_1^i(\bar{c}^i)]_2$  is the second coordinate of the vector of ratios of risk aversion. We start with endowments  $e^1 = (2 \ 3)^\top$  and  $e^2 = (6 \ 1)^\top$ . The two competitive traders' marginal utilities  $u_c^i(\cdot)$  are smooth functions as plotted in Fig 2.2.

Figure 2-2: Numerical Illustration With General Utility Functions



We can infer from Fig 2.2 that trader 1 and 2 have the same marginal utilities at consumption levels 2 and 4:  $u_c^1(2) = u_c^2(2) = 5$ ,  $u_c^1(4) = u_c^2(4) = 10$ . To exploit the fact that we can set the second derivative arbitrarily without violating any fundamental axiom of expected utility hypothesis, we let  $u_{cc}^1(2) = -50$ ,  $u_{cc}^2(2) = -25$ ,  $u_{cc}^1(4) = u_{cc}^2(4) = -100$ . This gives rise to an equilibrium with consumption allocation and price vector

$$\bar{c}_1^1 = \bar{c}_1^2 = 4, \bar{c}_2^1 = \bar{c}_2^2 = 2, \bar{\phi} = (1 \ 2)^\top$$

So in equilibrium, trader 1 sells 1 unit of consumption in state-2 in exchange for 2 unit of consumption in state-1:  $\bar{z}_1^1 = -\bar{z}_1^2 = 2$  and  $\bar{z}_2^1 = -\bar{z}_2^2 = -1$ . One can verify that

$$\alpha^1(\bar{c}_1^1) = \alpha^1(\bar{c}_2^1) = 10, \alpha^2(\bar{c}_1^2) = 5, \alpha^2(\bar{c}_2^2) = 10$$

and the expression in (2.9) is negative. We have  $\frac{\partial z_1}{\partial \phi_1} > 0$ . This implies  $\partial z$  is not negative semi-definite and there exists pure manipulation opportunities. In addition, we can also compute

$$\frac{\partial z_1^1(\bar{\phi})}{\partial \phi_1} = -\frac{13}{15}, \frac{\partial z_1^2(\bar{\phi})}{\partial \phi_1} = \frac{9}{10}$$

So agent 2 causes the upward-sloping demand. This shows, with general preferences, local pure manipulation could easily arise.

### CRRA Utilities

In this example, we shall consider competitive traders with CRRA utilities:  $u^i(c) = \frac{c^{1-\gamma^i}}{1-\gamma^i}$  where  $\gamma^i$  is trader  $i$ 's relative risk aversion. Recall that  $c^i(\phi) = \zeta^i(\phi, \phi^\top e^i)$  where  $\zeta^i(\phi, w^i)$  is the optimal consumption under state price vector  $\phi$  and wealth level  $w^i$ . For an endowments economy,  $w^i = \phi^\top e^i$ . It is well-known that CRRA utilities exhibits homotheticity<sup>16</sup>, which implies linear scaling in wealth:

$$\zeta^i(\phi, w^i) = w^i h^i(\phi) \quad (2.9)$$

where  $h^i(\phi) = \zeta^i(\phi, 1)$  is trader  $i$ 's unit wealth consumption demand. With wealth  $w^i$ , trader  $i$  will simply scale up her optimal consumption under unit wealth by a factor of  $w^i$ . CRRA utilities allow us to characterize  $h^i(\phi)$  easily and decompose  $\partial z^i$  into substitution and wealth effects as in (2.4).

**Proposition 12** *With CRRA utility function of the form  $u^i(c) = \frac{c^{1-\gamma^i}}{1-\gamma^i}$ , the unit wealth consumption demand of trader  $i$  is*

$$h^i(\phi) = \frac{\pi^{-\frac{1}{\gamma^i}}}{\phi^\top \pi^{-\frac{1}{\gamma^i}}} \quad (2.10)$$

where  $\pi = \left( \frac{\phi_0}{p_0} \frac{\phi_1}{p_1} \dots \frac{\phi_K}{p_K} \right)^\top$  is the state-price density vector<sup>17</sup>. Moreover,

$$\partial z^i = \underbrace{w^i s^i(\phi)}_{\text{Substitution}} \underbrace{- h^i(\phi) z^i(\phi)^\top}_{\text{Wealth Effect}} \quad (2.11)$$

where  $w^i = \phi^\top e^i$  is trader  $i$ 's wealth under  $\phi$  and  $s^i(\phi)$  is her Slutsky's substitution matrix with unit wealth

$$s^i(\phi) = -\frac{1}{\gamma^i} \left[ \begin{pmatrix} h_0^i(\phi)/\phi_0 & \dots & \dots & 0 \\ 0 & h_1^i(\phi)/\phi_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & h_K^i(\phi)/\phi_K \end{pmatrix} - h^i(\phi) h^i(\phi)^\top \right]$$

From the proposition above, we see the unit wealth consumption demand  $h^i$  is proportional to  $\pi^{-\frac{1}{\gamma^i}}$  and, as a result, the actual consumption  $c^i = w^i h^i$ , which is the unit wealth consumption scaled by trader  $i$ 's wealth. Therefore, the larger the variation of the state-price density across states, the higher the variation of consumption across states.

<sup>16</sup> A utility function is homothetic if  $U(\lambda c) = \lambda^k U(c)$  for some  $k > 0$ .

<sup>17</sup> We denote  $\pi^{-\frac{1}{\gamma^i}}$  as the coordinate-wise exponential  $\left( \pi_0^{-\frac{1}{\gamma^i}} \pi_1^{-\frac{1}{\gamma^i}} \dots \pi_K^{-\frac{1}{\gamma^i}} \right)^\top$ .

Moreover, trader with higher risk-aversion coefficient  $\gamma^i$  has consumption demand less sensitive to the variation of state-price density. This is because she is very unwilling to suffer low consumption in any state and aggressively smoothes consumptions across states

We also have a tractable decomposition of the derivative of excess consumption demand  $\partial z^i$ . The first component in (2.11) is the Slutsky's matrix captures the substitution effects. It measures the change in her consumption demand with respect to change in state prices if her wealth were compensated in a manner that would make her consumption demand before the price change just feasible. It is well-known from classical demand theory that the substitution effect is always negative semi-definite. With CRRA utility function, this component is simply the substitution effect under unit wealth  $s^i(\phi)$  scaled by the wealth level  $w^i$ . The higher the wealth level, the more pronounced the substitution effect, which leads to more negative semi-definite  $\partial z^i$ . To generate non-negative semi-definite  $\partial z^i$ , we can only rely on the second component  $-h^i z^{i\top}$ , which is due to wealth change caused by the price change. Its interpretation is intuitive: the  $j$ 's column is  $-z_j^i h^i$ , which measures the change in consumption across states due to a unit increase in state- $j$  price  $\phi_j$ . Again, the scaling property of CRRA utility implies that  $\partial_w \zeta^i(\phi, w^i) = h^i(\phi)$ : a unit change in wealth (keeping price fixed) will always cause the consumption to move by the unit wealth consumption  $h^i(\phi)$ . To undo the compensation in wealth included in the first component and take into consideration of change in market value of endowment when state price  $\phi_j$  increases by 1 unit marginally, a reduction in wealth by  $c^i - e^i = z^i$  is needed. Therefore, the combined effect of a marginal increase in  $\phi_j$  on consumption demand is  $-z_j^i h^i$ .

To fix ideas, let us assume there are two traders  $i = 1, 2$  with  $\gamma^1 > \gamma^2$ <sup>18</sup> and two states  $k = 1, 2$  at time-1. This is illustrated by Fig 2.3. Equilibrium state prices  $\bar{\phi}$  determines the gradient of the unit wealth budget line. Suppose that  $\phi_1 < \phi_2$  and  $\pi_1 < \pi_2$ , which could possibly be caused by low aggregate endowment in state-2. These are reflected by the flat budget line and the fact that  $h^1$  and  $h^2$  are all tilted towards state-1 consumption axis. As discussed above,  $h^2$  will respond to variation in state-price density  $\pi$  more sensitively since trader 2 has lower risk aversion. Therefore, as shown in Fig 2.3,  $h^2$  tilts towards more towards state-1 consumption axis than  $h^1$ . In equilibrium,  $z^1(\bar{\phi}) + z^2(\bar{\phi}) = 0$ , which implies

$$\begin{aligned} \partial z(\bar{\phi}) &= \partial z^1(\bar{\phi}) + \partial z^2(\bar{\phi}) \\ &= w^1 s^1(\bar{\phi}) + w^2 s^2(\bar{\phi}) - h^1(\bar{\phi}) z^1(\bar{\phi})^\top - h^2(\bar{\phi}) z^2(\bar{\phi})^\top \\ &= [w^1 s^1(\bar{\phi}) + w^2 s^2(\bar{\phi})] - [h^1(\bar{\phi}) - h^2(\bar{\phi})] z^1(\bar{\phi})^\top \end{aligned} \quad (2.12)$$

To have pure manipulation, we need  $\partial z(\bar{\phi})$  to be non-negative semi-definite. This means that  $v^\top \partial z^i v > 0$  for some unit vector  $v \in \mathbb{R}^2$ . We have already known that the substitution

<sup>18</sup>As shown in Proposition 10.1, if  $\gamma^1 = \gamma^2$ , a representative consumer exists and WARP holds at  $\bar{\phi}$ . To have local pure manipulation, we have to have different  $\gamma^1 \neq \gamma^2$ . W.l.o.g, let us assume  $\gamma^1 > \gamma^2$ .

effect is always negative:  $v^\top [w^1 s^1 + w^2 s^2] v < 0$ . The aggregate wealth effect is

$$\begin{aligned} v^\top [-(h^1 - h^2) z^{1\top}] v &= -[v^\top (h^1 - h^2)] (v^\top z^1) \\ &= -\|h^1 - h^2\|_2 \|z^1\|_2 \cos \vartheta_{vz} \cos \vartheta_{vh} \end{aligned}$$

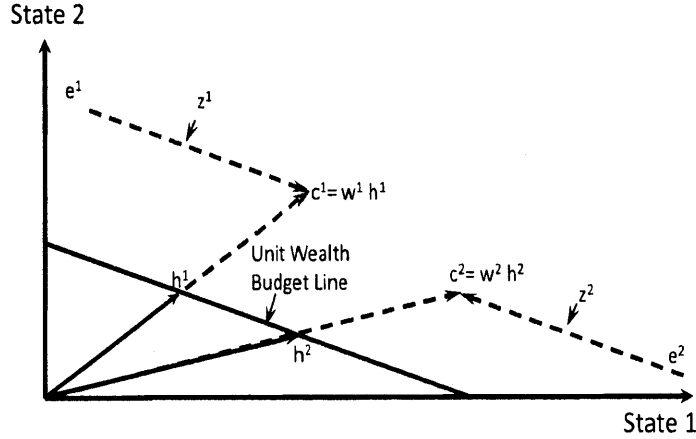
where  $\vartheta_{vz}$  and  $\vartheta_{vh}$  are angles between  $v$  and  $z^1$  as well between  $v$  and  $h^1 - h^2$ . To have large overall positive wealth, we need large  $\|h^1 - h^2\|_2$  and  $\|z^1\|_2$ . Furthermore, we need  $z^1$  and  $h^1 - h^2$  to be in opposite direction so that  $-\cos \vartheta_{vz} \cos \vartheta_{vh} > 0$ <sup>19</sup>. To have large  $\|h^1 - h^2\|_2$ , we need (1) the relative risk aversions  $\gamma^1$  and  $\gamma^2$  to be sufficiently different so that  $h^1$  and  $h^2$  are sufficiently different despite the same rankings of consumption levels across states and (2) variation in state price density  $\pi$  is large so that difference in risk aversions matters<sup>20</sup>. Moreover, we need (3) the trade amount  $\|z^1\|_2$  to be large. Additional, we must ensure that the direction of trade along the budget line by trader 1 is opposite to the difference in unit wealth consumption demands  $h^1 - h^2$ . Since competitive equilibrium with traders having CRRA utility functions allow linear scaling, we can fix the scale of trade  $z^1 = -z^2$ . To have minimal effect of the substitution effect  $[w^1 s^1(\bar{\phi}) + w^2 s^2(\bar{\phi})]$ , which prevents  $\partial z$  from becoming non-negative semi-definite, we would want to give traders as little wealth  $w^i$  as possible under a fixed equilibrium price  $\bar{\phi}$ . As we shrink  $w^i$ , consumption  $c^i = w^i h^i$  will scale back along  $h^i$  as illustrated in Fig 2.3. With  $z^i$  fixed, we are constrained by the requirement  $e^i = c^i - z^i \in \mathbb{R}_{++}^{K+1}$ . Thus, the substitution effect is minimized when either  $e^1$  or  $e^2$  has a coordinate that is about to hit zero. As shown by Fig 2.3, this leads to very different endowments with  $e^1$  rich in state-2 consumption but poor in state-1 consumption and  $e^2$  diametrically opposite. Thus, we need (4) heterogeneous endowments among traders. In summary, (1)-(4) corresponds the conditions we analyzed earlier on: heterogeneity among traders (both preferences  $\gamma^i$ 's and endowments  $e^i$ ), large trading needs (large  $\|z^1\|_2$ ) and large remaining insurance demand ( $\pi$  with high variations across states). These give rise to non-negative semi-definite  $\partial z(\bar{\phi})$ , which allows local pure manipulation.

The above intuition generalizes to more realistic cases with more than two states of nature. We shall demonstrate using a concrete case with time-0 consumption (state-0) and 3 states at time-1 (state-1, state-2, state-3). In order to have a volatile state price density  $\pi$ , we let the probability vector be  $p = (1 \ 0.001 \ 0.499 \ 0.5)^\top$ . Furthermore, let the equilibrium state prices be  $\bar{\phi} = (1 \ 40 \ 0.52 \ 0.48)^\top$ . Trader 1 is more risk-averse with  $\gamma^1 = 6$ . She has endowment  $e^1 = (0.0001 \ 0.0404 \ 0.0001 \ 0.0001)^\top$ . Trader 2 is less risk-averse with  $\gamma^2 = 3$ . She has endowment  $e^2 = (0.4981 \ 0.0001 \ 0.4926 \ 0.5037)^\top$ . In terms of aggregate endowment, state-1 is a "disaster" state, which happens with a rare probability 0.1%. State-2 is a "bad" state with aggregate endowment falling by 1.1% relative to state-0 aggregate endowment and state-3 is a "good" state with aggregate endowment growing by 1.1%. State-2 and state-3 happen with about equal chances. The state-price density at state-1 is very high due to its "disaster" nature. Trader 1 has almost 100% share of state-1 aggregate endowment but is poor in other states. Trader 2 is exactly in the opposite situation.

<sup>19</sup>In a 2-states case,  $z^1$  and  $h^1 - h^2$  are always proportional.

<sup>20</sup>If  $\pi \propto (1 \dots 1)^\top$ , then  $\gamma^1 \neq \gamma^2$  cannot lead to different  $h^i$ 's.

Figure 2-3: CRRA Utilities



She has almost 100% share of aggregate endowment in all other states except state-1, in which she has little endowment. In the competitive equilibrium without  $L$ , they will trade to share risk. The equilibrium price is  $\bar{\phi}$  as stated above and the consumption allocation is  $c^1 = (0.1827 \ 0.0312 \ 0.1814 \ 0.1839)^\top$  and  $c^2 = (0.3155 \ 0.0092 \ 0.3112 \ 0.3199)^\top$ . The trade amount is  $z^1 = -z^2 = (0.1826 \ -0.0091 \ 0.1813 \ 0.1838)^\top$ . Essentially, the two competitive traders trade to consume more in states, in which they are poorly endowed. The unit wealth consumption demands are  $h^1 = (0.1131 \ 0.0193 \ 0.1124 \ 0.1139)^\top$  and  $h^2 = (0.3155 \ 0.0092 \ 0.3112 \ 0.3199)^\top$  respectively. We can numerically verify that  $\partial z(\bar{\phi})$  is not negative semi-definite with a positive eigenvalue of 0.015 and the associated eigenvector is  $(1 \ -0.0382 \ 0.5463 \ 0.5076)^\top$ , which is the portfolio that is most upward-sloping in demand. This portfolio is very much similar to the trade portfolio between the two competitive traders. In this example, the demand curve for the "disaster" state consumption is also upward-sloping. However, we can construct many examples with downward-sloping demand curve for all individual state consumption but upward-sloping for a certain portfolio of securities. For instance, this can be achieved by setting  $e^1 = (0.0001 \ 0.0676 \ 0.0001 \ 0.6642)^\top$  and  $e^2 = (0.6575 \ 0.0001 \ 0.6509 \ 0.0001)^\top$ .

The economic intuition is as follows. We have a "Peso problem" that involves a disaster state in aggregate endowments. Trader 1 is more risk-averse and her endowment is rich in the disaster state but poor in other normal states. Trader 2 is less risk-averse and her endowment is exactly opposite to trader 1's. The difference in preferences and endowments lead to large trading needs in the competitive equilibrium without  $L$ . Moreover, the disaster state in aggregate endowments leaves large remaining insurance demand for the disaster state even after trading. If we reduce the normal state prices  $\phi_0$ ,  $\phi_2$  and  $\phi_3$  relative to the disaster state price  $\phi_1$ , trader 1 is relatively richer and trader 2 is relatively poorer. Ignoring the substitution effect for the moment, the wealth effect causes trader 2 to reduce consumptions along  $h^2$  in every states especially the normal ones. This is because trader 2's low risk-aversion leads her to respond  $\pi$  sensitively and she consumes

much more in the normal states that are "cheap". Trader 1 gains in wealth and consume more in every states along  $h^1$ . However, her high risk-aversion and, therefore, strong desire to smooth consumption across states leads her to spend much of her incremental wealth in the disaster state consumption, which is "expensive", and less in the normal states consumption. The reduction in normal states consumption from trader 2 outweighs the increase from trader 1. Therefore, the overall wealth effect is a reduction in normal states consumption demand. Substitution effect is linear in wealth  $w^1$  and  $w^2$  whereas the wealth effect is linear in the trading needs  $\|z^1\|_2$ . With wealth minimized holding trading needs fixed, the substitution effect, which is opposite to wealth effect here, is subdued. Overall, we will see a reduction in demand for consumptions in normal states if state prices in normal states are reduced relatively. This gives rise to pure manipulation opportunities.

### B. Pure Manipulation With Price-Dependent Beliefs

In the preceding analysis, we have studied pure manipulation that arises due to strong risk-sharing needs, heterogeneous preferences and endowments and remaining insurance demand even after substantial trading. Most importantly, our mechanism does not require any information advantage possessed by the large trader about either the underlying fundamental or her identity, which has been the focus of many previous studies of manipulation. A key element of these earlier manipulation mechanisms is that competitive traders' belief about the probability distribution of states of nature is dependent on security prices they observe as a result of actual or potential informational inferiority of the competitive traders. Here we shall show that this element could also be captured in our framework. For simplicity, we shall maintain the assumption of complete markets and, in addition, assume that competitive traders are homogeneous in preference and endowments. While we have demonstrated that pure manipulation is not possible against homogeneous competitive traders earlier on, the incorporation of price-dependent belief opens a new channel for price manipulation to affect competitive traders' excess consumption demand function and allows pure manipulation to exist.

We assume a general form of probability distribution dependency on prices. In particular, let the probability of state- $k$  perceived by competitive traders be  $\tilde{p}_k(\phi, \bar{\phi}, p)$ , which is a function of the actual probability distribution  $p$ , the state price vector  $\bar{\phi}$  in the competitive equilibrium as well as the new state price vector  $\phi$  set by the large trader. The interpretation of this functional form is straightforward. Competitive traders start with the actual probability distribution  $p$ . They observe the change of the state price vector  $\phi$  relative to the competitive equilibrium one  $\bar{\phi}$  brought by the large trader and update their belief to  $\tilde{p}$ . They update because they believe the large trader may have a different probability distribution, which is more accurate than their own. The functional form  $\tilde{p}_k(\phi, \bar{\phi}, p)$  is left to be specified exogenously. It could be interpreted as a behavioral assumption or a reduced form of equilibrium results. Of course, we must impose that  $\tilde{p}$  satisfies:

$$\tilde{p}_0(\phi, \bar{\phi}, p) = 1, \sum_{k=1}^K \tilde{p}_k(\phi, \bar{\phi}, p) = 1 \text{ and } \tilde{p}_k(\phi, \bar{\phi}, p) > 0 \quad (2.13)$$

which are basic conditions for an equivalent probability measure. With homogeneous

competitive traders, we know that the competitive equilibrium without  $L$  is unique. We shall fix  $\bar{\phi}_0 = 1$  to avoid multiple equilibrium nominal state price vectors. Besides, since  $\tilde{p}$  should be homogeneous in  $\phi$  of degree 0 as scaling all state prices by the same constant should not have any actual effect:

$$\tilde{p}_k(\lambda\phi, \bar{\phi}, p) = \tilde{p}(\phi, \bar{\phi}, p) \forall \lambda > 0 \quad (2.14)$$

To isolate the effect of large trader manipulation, we assume that, under the state price vector of competitive equilibrium without  $L$ , competitive traders' probability distribution vector is the actual one:

$$\tilde{p}(\bar{\phi}, \bar{\phi}, p) = p \quad (2.15)$$

Again, to have local pure manipulation, we need  $\partial z$  to be non-negative semi-definite. So the following proposition gives an explicit characterization of  $\partial z(\bar{\phi})$ . With homogeneous competitive traders, we may assume that there is a single representative competitive trader and drop the index  $i$ .

**Proposition 13** *In competitive equilibrium without  $L$  (or price-taking  $L$ ),*

$$\begin{aligned} \frac{\partial c_k(\bar{\phi})}{\partial \phi_j} &= \frac{1}{\alpha(e_k)} \left[ \frac{b_j}{\bar{\phi}_j} + \sum_{l=0, l \neq k}^K b_l \frac{\partial \ln(\tilde{p}_k/\tilde{p}_l)}{\partial \phi_j} \right], j \neq k \\ \frac{\partial c_k(\bar{\phi})}{\partial \phi_k} &= \frac{1}{\alpha(e_k)} \left[ \sum_{l=0, l \neq k}^K b_l \left( \frac{\partial \ln(\tilde{p}_k/\tilde{p}_l)}{\partial \phi_k} - \frac{1}{\bar{\phi}_k} \right) \right], \text{ otherwise} \end{aligned}$$

where  $b_l = \frac{\bar{\phi}_l}{\alpha(e_l)} / \sum_{j=0}^K \frac{\bar{\phi}_j}{\alpha(e_j)}$ .

Similar to Corollary 2, we have local pure manipulation if the excess consumption demand is upward-sloping for a state.

**Corollary 4** *There is local pure manipulation if*

$$\max_{b \in \{0,1,\dots,K\}} \left( \sum_{l=0, l \neq k}^K b_l \frac{\partial \ln(\tilde{p}_k/\tilde{p}_l)}{\partial \phi_k} \right) - \frac{1 - b_k}{\bar{\phi}_k} > 0$$

For a solid example, consider the following updating rule for competitive traders

$$\tilde{p}_k(\phi, \bar{\phi}, p) = \frac{\max \left[ \left( \frac{\phi_k}{\bar{\phi}_k} \right)^\beta, \varepsilon \right] p_k}{\sum_{l=1}^K p_l \max \left[ \left( \frac{\phi_l}{\bar{\phi}_l} \right)^\beta, \varepsilon \right]} \text{ for some } \beta > 0, 0 < \varepsilon \ll 1$$

Clearly, this satisfies our assumptions (2.13), (2.14) and (2.15). A natural interpretation is that if competitive traders see a state price  $\phi_k$  set to be higher than  $\bar{\phi}_k$ , they believe

that the large trader thinks state- $k$  is more likely to happen than the actual probability  $p_k$  and is willing to pay a higher state price  $\phi_k$  in order to induce competitive traders to sell. If they also believe that the larger trader's probability distribution is more accurate than theirs, they will update accordingly. The sensitivity of updating is determined by  $\beta$ . Larger  $\beta$  leads competitive traders to place more probability on state- $k$  when they see the large trader sets  $\phi_k > \bar{\phi}_k$ . Conversely, if  $\phi_k < \bar{\phi}_k$ , competitive traders will reduce probability on state- $k$  until hitting a small lower bound  $\varepsilon$ , which is introduced to prevent zero probability. One can easily verify that  $\frac{\partial z_k(\bar{\phi})}{\partial \phi_k} = (\beta - 1) / \bar{\phi}_k$ . Thus, if  $\beta > 1$ , we will have locally upward-sloping demand curve for each state at  $\bar{\phi}$ , which gives rise to local pure manipulation.

## 2.3 Extension To Multi-Period Economy

In this section, we shall extend our setup to a multi-period economy. Interestingly, although price manipulation has most often been considered under a multi-period dynamic setup, we find that being dynamic adds nothing to the large trader's price manipulation power. In fact, it only reduces her price manipulation power due to the typical commitment problem faced by an intertemporal monopolist. This interesting observation suggests we should consider manipulation by a large trader in a static setup, the success of which is necessary for manipulation in a dynamic setup. Moreover, although it is often thought that different security structures do not matter for allocation in competitive equilibrium when market is complete, they do make a difference with a large trader present. We shall also outline a condition that allows pure manipulation to exist for a multi-period security structure that is in the same vein as Proposition 8.

### 2.3.1 Model Setup

The information revelation process we shall consider is a typical event-tree structure with  $t = 0, 1, \dots, T$ . For brevity, we shall focus on the notations rather than a rigorous formulation. A node in the tree is denoted as  $\xi$ .  $\xi_0$  is the time-0 state and the root of the tree. Let  $\xi^+$  denote the set of immediate offspring nodes of  $\xi$ . Furthermore, let  $\mathbb{D}(\xi)$  be the set of nodes that are the descendants of node  $\xi$  including  $\xi$  itself. Hence,  $\mathbb{D}(\xi)$  is the subtree originating from node  $\xi$ <sup>21</sup>. We say  $\xi$  precedes  $\xi'$  if  $\xi' \in \mathbb{D}(\xi)$ . A node is associated with a time of occurrence, which we denote as  $t(\xi)$ . With this, we can define the set of nodes occurring at time- $\tau$   $\Omega_\tau = \{\xi : t(\xi) = \tau\}$  for  $\tau = 0, 1, \dots, T$ . Thus, the set of terminal nodes  $\Omega_T$  is the state space. As the subsequent analysis involves collapsing the event-tree into a sequence of nodes, we shall define an order of traversal here so that the  $k$ th visited node will correspond to the  $k$ th element in the sequence. First, we shall assign an order  $1, 2, \dots, |\xi^+|$ <sup>22</sup> to the offspring nodes of  $\xi$  if  $|\xi^+| > 0$ . This allows us to draw

<sup>21</sup>Therefore  $\mathbb{D}(\xi_0)$  is the whole event-tree.

<sup>22</sup>We shall denote the number of elements in a set  $A$  as  $|A|$ .



the event-tree with time increases from left to right. Our order of traversal will be to go down the column of nodes of time- $t$  from top to bottom and then move on to the next column of nodes of time- $t + 1$ .

Let  $\mathbb{P}$  be a probability measure defined on  $\Omega_T$  and  $\mathbb{F}_0 \subset \mathbb{F}_1 \subset \dots \subset \mathbb{F}_T$  be the natural filtration associated with the event-tree. We shall write  $p(\xi) = \sum_{\omega \in \Omega_T \cap \mathbb{D}(\xi)} \mathbb{P}(\{\omega\})$  as the probability of occurrence of node  $\xi$ . Furthermore, we shall write the probability of occurrence of node  $\xi'$  conditional on reaching a node  $\xi$  preceding  $\xi'$  as  $p^\xi(\xi') = p_{\xi'}/p_\xi$  where  $\xi' \in \mathbb{D}(\xi)$ .

Traders  $i = 1, 2, \dots, N, L$  have time-additive von Neumann-Morgenstern utility  $U^i(c^i) = \sum_{\xi \in \mathbb{D}(\xi_0)} p(\xi) u^i(c^i(\xi))$  where  $c^i(\xi)$  is the consumption at node  $\xi$  and  $u^i$  is the Bernoulli utility function satisfying the earlier assumptions. The endowment of trader  $i$  at node  $\xi$  before any trading is  $e^i(\xi)$ . After each trading, trader  $i$ 's endowment at node  $\xi$  may be changed. With some abuse of notation, let us write the endowment of trader  $i$  at node  $\xi'$ , just before trading at a preceding node  $\xi$  takes place, as  $e^i(\xi, \xi')$  where  $\xi' \in \mathbb{D}(\xi)$ . Therefore, the original endowment without any trading is  $e^i(\xi_0, \xi)$ .

There are  $M$  securities (security  $1, 2, \dots, M$ ) in total. They could be long-lived or short-lived. Security  $m$  begins trading at time  $0 \leq \tau_m^{start} < T$  and vanishes at time  $0 < \tau_m^{end} \leq T$ . Security  $m$  pays dividend  $\delta_m(\xi)$  at node  $\xi$  if  $\tau_m^{start} < t(\xi) \leq \tau_m^{end}$ . Let the set of tradable assets at node  $\xi$  be  $\mathbb{M}(\xi)$ . We do not need to assume that security  $m$  must be traded at every period between  $\tau_m^{start}$  and  $\tau_m^{end}$ <sup>23</sup>. Suppose security  $m$  is traded at node  $\xi$ . Let  $\eta(m, \xi)$  denote the node, at which security  $m$  is tradable last time. The security prices are determined endogenously. The price of security  $m$  at node  $\xi$  is  $S_m(\xi)$  (if  $m \in \mathbb{M}(\xi)$ ). Also, let vector  $S(\xi)$  of dimension  $|\mathbb{M}(\xi)| \times 1$  denote the prices of securities tradable at node  $\xi$ . We can summarize the prices in a vector  $S$  of dimension  $\sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)| \times 1$ . Each coordinate in  $S$  corresponds to the price of security  $m$  at node  $\xi$  ( $m \in \mathbb{M}(\xi)$ ). Thus, we can use the pair  $(m, \xi)$  to refer to the coordinate position. Coordinate  $(m, \xi)$  is before  $(m', \xi')$  if  $\xi$  is before  $\xi'$  in the event-tree traversal order or, when  $\xi = \xi'$ ,  $m < m'$ . The payoff matrix  $X(S)$  is a matrix of dimension  $|\mathbb{D}(\xi_0)| \times \sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)|$ . Each column corresponds to the consumption payoff at all nodes of buying 1 unit of security  $m$  available for trade at node  $\xi$  (i.e.  $m \in \mathbb{M}(\xi)$ ) and holding it until the next node where  $m$  is tradable again (if there is a node  $\xi' \in \mathbb{D}(\xi)$  and  $m \in \mathbb{M}(\xi')$ ). These columns are arranged in the same order as coordinates of  $S$ . Along each column, the coordinates are arranged in the same order as event-tree traversal.

The security demand of competitive trader  $i$  is denoted as  $\theta^i$  of dimension  $\sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)| \times 1$ <sup>24</sup> where each coordinate  $(m, \xi)$  of  $\theta^i$  represents buying  $\theta_m^i(\xi)$  units of a security  $m$  at node  $\xi$  and holding them until the next node  $\xi' \in \mathbb{M}(\xi)$  where  $m$  becomes tradable. The large trader has price-setting power but has to clear the market by absorbing the competitive traders' aggregate security demand  $\theta^L(S) = -\sum_{i=1}^N \theta^i(S)$ .

<sup>23</sup>That is to say  $\tau_m^{start} \leq t(\xi) < \tau_m^{end}$  does not imply  $m \in \mathbb{M}(\xi)$ .

<sup>24</sup>Coordinates are arranged in the same order as  $S$ .

There are two large trader equilibrium concepts for a multi-period economy. In a **commitment equilibrium**, the large trader announces a security price vector  $S$  at time-0 and is able to commit to set prices for tradable securities  $\mathbb{M}(\xi)$  at each node  $\xi$  according to  $S$  over time while competitive traders choose security demand  $\theta^i(S)$ 's to maximize utilities. The equilibrium price  $\hat{S}^{Com}$  maximizes the large trader's utility. In a **dynamic equilibrium**, when a node  $\xi$  is reached, the large trader sets prices of security available for trade  $S(\xi)$  while competitive traders take  $S(\xi)$  as given, conjecture about security prices at subsequent nodes of  $\xi$ ,  $S^e(\xi, \xi') \forall \xi' \in \mathbb{D}(\xi) \setminus \xi$ , and form optimal security demand  $\theta^i(\xi, S(\xi))$ 's. For equilibrium price vector  $\hat{S}^{Dyn}$ , we require  $\hat{S}^{Dyn}$  to maximize utility for the large trader taking competitive traders' conjecture as given and competitive traders are fully rational:  $S^e(\xi, \xi') = \hat{S}^{Dyn}(\xi') \forall \xi, \xi'$  such that  $\xi' \in \mathbb{D}(\xi)$ .

### 2.3.2 Security Structures

In a commitment equilibrium, given a security structure  $X(S)$ , the large trader can choose any security price vector  $S$  as long as it does not cause any arbitrage. However, she does not enjoy the same freedom in a dynamic equilibrium. The large trader's commitment equilibrium price-setting plan becomes time-inconsistent as remarked in Basak (1997) in the sense that she has an incentive to deviate from the time-0 price-setting plan when a later node  $\xi$  is reached, which is similar to the durable-good monopolist's problem. In other words, if competitive traders believed that large trader would follow the optimal commitment equilibrium price-setting plan and trade accordingly, the large trader would choose different prices  $S(\xi)$  at  $\xi$  since the endowments of competitive traders  $e^i(\xi, \xi')$ 's are already changed after the initial trading. Thus, this leads us to the following proposition.

**Proposition 14** *Given a security structure  $X(S)$ , the large trader yields higher or equal utility in a commitment equilibrium than in a dynamic equilibrium.*

**Proof.** Consider a dynamic equilibrium. Full rationality requires that competitive traders' expectation about future prices never change as time evolves and are the same as what the large trader actually sets. Thus, effectively, the competitive traders' security demand  $\theta^i$  would be the same as that in a commitment equilibrium with  $\hat{S}^{Dyn}$  being the chosen security price vector. But since  $\hat{S}^{Dyn}$  is not necessarily the optimal one in a commitment equilibrium,  $U^L(e^L + X(\hat{S}^{Dyn})\hat{\theta}^{Dyn}) \leq U^L(e^L + X(\hat{S}^{Com})\hat{\theta}^{Com})$ . ■

The proposition above leads us to consider whether security structures that give rise to the same asset span are the same in a dynamic equilibrium. For simplicity, let us consider complete markets. Markets can be complete with contingent claims traded only at time-0. Since there is no further trading beyond time-0, this allows the large trader to commit. Thus, the equilibrium would be the same as a commitment equilibrium. However, there are many other ways to complete the markets. We can have sufficient number of long-lived securities<sup>25</sup> traded every period. Alternatively, we can also have

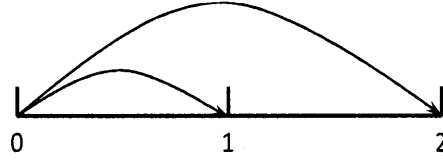
<sup>25</sup>By "long-lived", we mean the security will not vanish until terminal time  $T$ .

short-lived securities<sup>26</sup> traded every period. When all traders behave competitively, it is well-known that complete markets yield the same equilibrium allocation regardless the underlying security structures. However, this is no longer the case for a large trader dynamic equilibrium. Trading at later times causes the commitment problem for the large trader. As a result, trading only contingent claims at time-0 will yield higher utility for the large trader than other security structures that complete the market dynamically. The equilibrium allocations are, therefore, different. We shall illustrate with the following example.

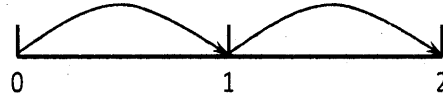
### 3-Period Economy With No Uncertainty

Consider a 3-period economy with  $t = 0, 1, 2$  with no uncertainty. Let us denote the three nodes in this simple event-tree as  $\xi_0, \xi_1$  and  $\xi_2$ . The large trader is risk-neutral and competitive traders have log-utility with homogeneous endowments  $e = (e(\xi_0) \ e(\xi_1) \ e(\xi_2))^T = (1 \ g \ g^2)^T$  where  $g > 1$  can be interpreted as the gross growth rate of the endowment stream. We shall study 4 security structures:

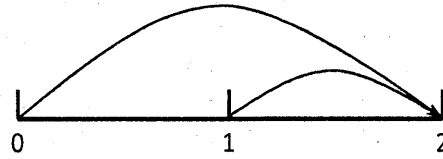
1. Contingent claims traded at time-0 only



2. Only short-lived bonds traded at time-0 and time-1



3. Long-lived zero-coupon bond traded at time-0 and re-traded at time-1



Since our focus is the intertemporal commitment problem for the large trader, we leave out uncertainties for simplicity. Under all three security structures, we have complete markets. However, the security structures are different.

In security structure 1, we have contingent claims traded at time-0 only. As there is no re trading later on, this will deliver the same results as a commitment equilibrium. Therefore, let the associated state price vector be  $\phi = (1 \ \phi_1(\xi_0) \ \phi_2(\xi_0))^T$ . The large

<sup>26</sup>By "short-lived", we mean the security will exist for only one period.

trader sets  $\phi$ . Competitive traders solve

$$\max_c \log c(\xi_0) + \log c(\xi_1) + \log c(\xi_2) \text{ s.t } \phi^\top c \leq \phi^\top e \quad (2.16)$$

It is easy to show that  $c = \left( \frac{\phi^\top e}{3} \frac{\phi^\top e}{3\phi_1(\xi_0)} \frac{\phi^\top e}{3\phi_2(\xi_0)} \right)^\top$ . Since the large trader is risk-neutral, she sets  $\phi$  by solving

$$\max_\phi \sum_{t=0}^2 e(\xi_t) - c(\xi_t) \quad (2.17)$$

Upon solving, we have  $\hat{\phi} = \left( 1 \sqrt{\frac{e(\xi_0)}{e(\xi_1)}} \sqrt{\frac{e(\xi_0)}{e(\xi_2)}} \right)^\top = (1 \ g^{-1/2} \ g^{-1})^\top$ . Therefore,

$$\begin{aligned} \hat{c} &= \frac{1 + g^{1/2} + g}{3} (1 \ g^{1/2} \ g)^\top \\ \hat{z} &= \frac{1}{3} (g^{1/2} + g - 2 \ g^{1/2} + g^{3/2} - 2g \ g + g^{3/2} - 2g^2)^\top \end{aligned}$$

. Clearly,  $\hat{z}(\xi_0) > 0$  and  $\hat{z}(\xi_1) = g^{1/2}(g^{1/2} - 1)^2 > 0$  but  $\hat{z}(\xi_2) < 0$ . So competitive traders sell consumption at  $\xi_2$  for consumption at  $\xi_0$  and  $\xi_1$ . This is expected as their endowment is highest at  $\xi_2$ . And the large trader improves her utility by  $\frac{2}{3}(1 + g^2 - g^{1/2} - g^{3/2}) > 0$ <sup>27</sup>.

Under security structure 2, a short-lived bond is traded at time-0 to exchange consumption between  $\xi_0$  and  $\xi_1$ . Its price is  $\phi_1(\xi_0)$ . Another short-lived bond is traded at time-1 to exchange consumption between  $\xi_1$  and  $\xi_2$ . Its price is  $\phi_2(\xi_1)$ . As there are two rounds of trading, the large trader cannot credibly commit to the optimal price-setting plan. To solve for the equilibrium, we shall use backwards induction. At  $\xi_1$  (after trading at  $\xi_0$ ), the competitive traders endowments at  $\xi_1$  and  $\xi_2$  are  $e(\xi_1, \xi_1)$  and  $e(\xi_1, \xi_2)$  respectively. At  $\xi_1$ , competitive traders solve

$$\max_c \log c(\xi_1) + \log c(\xi_2) \text{ s.t } c(\xi_1) + \phi_2(\xi_1) c(\xi_2) \leq e(\xi_1, \xi_1) + e(\xi_1, \xi_2) \phi_2(\xi_1) \quad (2.18)$$

It can easily be shown that  $c(\xi_1) = \frac{e(\xi_1, \xi_1) + e(\xi_1, \xi_2) \phi_2(\xi_1)}{2}$  and  $c(\xi_2) = \frac{e(\xi_1, \xi_1) + e(\xi_1, \xi_2) \phi_2(\xi_1)}{2\phi_2(\xi_1)}$ . The large trader sets  $\phi(\xi_1)$  by solving

$$\max_{\phi_2(\xi_1)} \sum_{t=1}^2 e(\xi_t) - c(\xi_t) \quad (2.19)$$

which gives

$$\phi_2(\xi_1) = \sqrt{\frac{e(\xi_1, \xi_1)}{e(\xi_1, \xi_2)}} \quad (2.20)$$

---

<sup>27</sup>This can be seen by noting that  $g^x$  is a convex function in  $x$ .

Her continuation value function is

$$V^L(\xi_1) = e^L(\xi_1, \xi_1) + e^L(\xi_1, \xi_2) + \frac{1}{2} \left( \sqrt{e(\xi_1, \xi_1)} - \sqrt{e(\xi_1, \xi_2)} \right)^2.$$

Going back to  $\xi_0$ , the large trader sets the price of short-lived bond,  $\phi_1(\xi_0)$  that exchange consumption at  $\xi_0$  for consumption at  $\xi_1$ . Competitive traders have a expected price for the short-lived bond at  $\xi_1$ ,  $\phi^e(\xi_0, \xi_1)$  (abbreviated as  $\phi^e$ ). The effective state prices  $\phi = (1 \ \phi_1(\xi_0) \ \phi_1(\xi_0) \ \phi^e)^\top$ . They conduct the same optimization as (2.16) and their consumption demand is

$$c = \left( \frac{\phi^\top e}{3} \quad \frac{\phi^\top e}{3\phi_1(\xi_0)} \quad \frac{\phi^\top e}{3\phi_1(\xi_0)\phi^e} \right)^\top$$

This implies that they need to buy  $\frac{\phi^\top e}{3} - e(\xi_0)$  units of consumption at  $\xi_0$  and, in exchange, sell  $\left( \frac{\phi^\top e}{3} - e(\xi_0) \right) \frac{1}{\phi_1(\xi_0)}$  units of consumption at  $\xi_1$ . Hence, the after-trading endowment of competitive traders is  $e(\xi_1, \xi_1) = e(\xi_1) - \left( \frac{\phi^\top e}{3} - e(\xi_0) \right) \frac{1}{\phi_1(\xi_0)}$  and  $e(\xi_1, \xi_2) = e(\xi_2)$ . But since competitive traders' expectation is rational, we need  $\phi^e = \phi_2(\xi_1)$ . This implies, in equilibrium

$$\begin{aligned} \phi^e &= \sqrt{\frac{e(\xi_1, \xi_1)}{e(\xi_1, \xi_2)}} \\ &= \sqrt{\frac{e(\xi_1) - \left( \frac{\phi^\top e}{3} - e(\xi_0) \right) \frac{1}{\phi_1(\xi_0)}}{e(\xi_2)}} \end{aligned} \quad (2.21)$$

Large trader solves

$$\max_{\phi=(1 \ \phi_1(\xi_0) \ \phi_1(\xi_0) \ \phi^e)^\top} e(\xi_0) - c(\xi_0) + V^L(\xi_1) \quad (2.22)$$

taking  $\phi^e$  as given. Rational conjecture by competitive traders dictates that

$$\hat{\phi}^e = \sqrt{\frac{e(\xi_1) - \left( \frac{\phi^\top e}{3} - e(\xi_0) \right) \frac{1}{\phi_1(\xi_0)}}{e(\xi_2)}} \quad (2.23)$$

Notice that in contrast with (2.17), the large trader loses 1 degree of freedom since she cannot set  $\phi^e$  directly. The constraint arises because the large trader can only set the price for short-lived bond at  $\xi_0$  and is unable to commit to a price of the short-lived bond at  $\xi_1$ , which is left to competitive traders to conjecture. As the large trader is optimizing within a subset of choices allowed under security structure 1, we expect the utility of the large trader to be lower.

Under security structure 3, a long-lived bond is traded at  $\xi_0$  and again at  $\xi_1$ . The trade at  $\xi_0$  allows competitive traders to exchange consumption at  $\xi_0$  for consumption at  $\xi_2$ . The second trade at  $\xi_1$  allows them to exchange consumption at  $\xi_1$  for consumption at  $\xi_2$ .

By symmetry, the dynamic equilibrium would be the same as under security structure 2 with  $\xi_1$  and  $\xi_2$  swapped. Thus, we expect that the utility of the large trader is lower than under security structure 1 due to the commitment problem. The main purpose of trading for competitive traders is to smooth consumption intertemporally. In particular, they sell consumption units at  $\xi_2$  in exchange for consumption units at  $\xi_0$  since their endowment is highest in  $\xi_2$  and lowest in  $\xi_0$ . Security structure 3 does allow the large trader to commit to the ratio of state prices at  $\xi_0$  and  $\xi_2$  directly as oppose to security structure 2. This mitigates the commitment problem. We expect the large trader's utility is higher under security structure 3 than under security structure 2.

For a numerical Illustration, let  $g = 105\%$ . Under three different types of security structures, the large trader's profit, competitive traders' utility and effective state prices are presented in Table 2.2.

Table 2.2: Dynamic Equilibrium Under Various Complete Markets Security Structures

	Effective State Prices	$L$ 's Profit	Comp Traders' Utility
Security Structure 1	$(1 \ 0.976 \ 0.952)^T$	0.0013	0.1470
Security Structure 2	$(1 \ 0.985 \ 0.951)^T$	0.0012	0.1469
Security Structure 3	$(1 \ 0.980 \ 0.965)^T$	0.0012	0.1471

It is interesting that security structure 2 leads to a Pareto disimprovement relative to security structure 1. Although dynamic trading reduces manipulation power of the large trader under security structure 2, it does not necessary imply that competitive traders will fare better. In the above example, competitive traders' main trading needs is to tranfer consumption from  $\xi_2$  to  $\xi_0$ . Security structure 2 does not allow them to do so directly. The only viable way is to transfer consumption from  $\xi_1$  to  $\xi_0$  first and then from  $\xi_2$  to  $\xi_1$ . However, if the after-trading (at time-0) endowments at  $\xi_1$  and  $\xi_2$  are widely apart, competitive traders expect the large trader to set a high price for consumption at  $\xi_1$  ( $\phi^e$  will be low). As a result, they are not willing to subsidize consumption at  $\xi_0$  heavily knowing that they would have to transfer consumption from  $\xi_2$  to  $\xi_1$  at an unfavourable term. This leads to overall insufficient intertemporal consumption sharing between the large trader and the competitive traders and results in a Pareto disimprovement.

### 2.3.3 Pure Manipulation In Commitment Equilibrium

In this subsection, we shall consider pure manipulation in a multi-period economy. For simplicity, we consider the commitment equilibrium. If there is no pure manipulation in a commitment equilibrium, Proposition 14 suggests that there is no pure manipulation in a dynamic equilibrium either.

The notion of no trading motive in multi-period economy is the same as in the static economy: the large trader is introduced so that she would not trade in the prevailing competitive equilibrium without  $L$  if she were to behave competitively.

**Definition 7** *Suppose the security prices in the prevailing competitive equilibrium without  $L$  is  $\bar{S}$ . If  $\partial U^L(e^L) X(\bar{S}) = 0$ , there is no trading motive between  $L$  and the rest.*

The notion of pure manipulation is exactly the same as in the static economy: if the large trader has no trading motive with respect to the prevailing competitive equilibrium without  $L$  but still can improve her utility through price manipulation (locally if there are multiple equilibria), then it is considered as pure manipulation.

By the same token as Proposition 4, we have the following proposition.

**Proposition 15** *If  $\partial U^L(e^L) X(S) \theta(S) > 0$  for all  $S$ , there is no pure manipulation. Otherwise, there is pure manipulation with a sufficiently risk-tolerant large trader.*

Again, we have a sufficient condition for local pure manipulation to exist. Let  $D^L$  be a diagonal matrix of dimension  $\sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)| \times \sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)|$ . The vector of diagonal entries has the same order as  $S$  with the entry corresponding to security  $m$  tradable at node  $\xi$  being  $u_c^L(e^L(\xi))$ .  $D^L$  is a scaling matrix consisting of the large trader's marginal utility at each node. Also define  $B$  as the shift matrix of dimension  $\sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)| \times \sum_{\xi \in \mathbb{D}(\xi_0)} |\mathbb{M}(\xi)|$ . Consider security  $m$  tradeable at node  $\xi$ . If there is an immediate preceding node  $\xi' = \eta(m, \xi)$  such that  $m$  is traded, let the entry corresponding to  $[(m, \xi), (m, \xi')]$  of  $B$  be 1. All other entries of  $B$  are zero. We have a proposition similar to Proposition 8.

**Proposition 16** *If  $(I - B) D^L \partial \theta$  is NOT negative semi-definite, then there is local pure manipulation.*

When the large trader sets price for a security that has been traded before, not only does it indicate the amount of consumption a buyer needs to pay to gain the payoff delivered until the next tradable node (this is captured by  $I$ ), it also determines the capital gain for a buyer who bought this security at the previous tradable node. This re-trading effect is captured by  $B$ . Trading takes at different nodes and the large trader's marginal valuations of each unit of consumption at these nodes are different. Utility improvement brought about by a bundle of consumption at different nodes as a result of manipulating the price locally needs to take into account different marginal valuations, which is captured by the scaling matrix  $D^L$ . If we have a static economy or contingent claim traded only at  $\xi_0$ ,  $B = 0$  and  $D^L = u_c^L(e^L(\xi_0)) I$ . This allows us to go back to Proposition 8.

## 2.4 Conclusion

In this paper, we study price manipulation in general equilibrium with no other market imperfection than the presence of a large trader. We show that this simple setup still allows pure manipulation, which is malicious in nature, to exist. We also discover that an important criterion for pure manipulation to exist under a general security structure – failure of the Weak Axiom of Revealed Preference at the competitive equilibrium in the absence of the large trader. The failure is associated with heterogeneity in preferences and endowments, large trading needs and remaining insurance demand in the competitive equilibrium, which causes the joint security demand curve to be upward-sloping for a certain portfolio of securities. In comparison with mechanisms proposed in earlier studies, ours does not require any form of information asymmetry, behavioral biases among competitive traders or malfunctioning security markets. The only imperfection needed is being large and able to move prices. Moreover, it is worth noting that, our mechanism often requires only local price-setting ability and, therefore, small capital commitment. These two features suggest that our mechanisms might be more ubiquitous and harder to detect by the market regulator. Besides, we find that security structures matter in the presence of a large trader even if the effective asset spans are the same, which forms a stark contrast with the case with all traders behaving competitively. The main reason is that security structures that allow dynamic trading reduces manipulation power. Hence, more illiquid products like corporate bonds, exotic options or bespoke CDOs might be more susceptible to price manipulations.

There are a few interesting directions that are worth further exploration. In particular, we assume that competitive traders all have the same belief. Heterogeneity in beliefs could also lead to pure manipulation even if we remove certain factors such as heterogeneous preferences that are important to pure manipulation in our current analysis. Moreover, we do not feature a production side and endogenous security design in our analysis. In reality, a large trader could control the supplies of assets. Large mortgage issuers like Freddie Mac and Fannie Mae control both the supply of mortgage products and have significant pricing power. How the origination of securities interacts with pricing power is a question out of the scope of current analysis. Besides, it is often the case that security structures are not exogeneously given. Financial innovations are actively undertaken by investment banks who also possesses significant pricing power. As discussed earlier on, security structures matter in a multi-period economy. Therefore, it would be interesting to analyze how a large trader designs a security structure to maximize her manipulation power. Finally, as a caveat, we restrict to a single large trader in our study. While in certain markets of limited capitalization, such a dominant player may exist. More generally, there could multiple non-competitive traders. Their competition, in either Cournot or Bertrand form, should reduce manipulation power to a certain extent. How conditions for pure manipulation in the presence of multiple large traders differ from our current ones would be left for future research.



## Appendix

### Proof of Proposition 2:

Consider the set

$$\mathcal{V}^L = \{V^L(\theta^L(S)) > -\infty : S \in \mathbb{R}^{M+1}\}$$

Our assumptions about competitive traders' utility functions and endowments ensure that the "competitive equilibrium without  $L$ " exists (See Magill and Quinzii (1996) Chapter 2 for details). By setting  $S$  equal to  $\bar{S}$ , the market-clearing price of the "competitive equilibrium without  $L$ ", the aggregate security demand  $\theta(\bar{S}) = 0$ . Since we have assumed that the large trader can survive without having to trade with the competitive traders (i.e.  $U^L(e^L) > -\infty$ ),  $\bar{S} \in \mathcal{V}^L$ . So  $\mathcal{V}^L$  is non-empty. By the non-negative consumption condition for competitive traders, trading will result in at most all the endowments of competitive traders being transferred to the large trader. Hence,  $\mathcal{V}^L$  is bounded above by  $U^L(e^L + e)$ . Since  $\mathbb{R}$  is complete, every non-empty set bounded above has a supremum. Let  $\hat{V}^L$  denote the supremum of set  $\mathcal{V}^L$ . Hence, there exists a price sequence  $S^{(1)}, S^{(2)}, \dots, S^{(n)} \dots$  such that  $V^L(\theta^L(S^{(n)}))$  is strictly increasing and  $V^L(\theta^L(\bar{S})) < V^L(\theta^L(S^{(n)})) \rightarrow \hat{V}^L$  as  $n \rightarrow \infty$ .

Since security prices are nominal, we can normalize all non-zero price vector  $S$  by its Euclidean norm and consider its equivalent price vector in

$$\mathcal{S} = \{S \in \mathbb{R}^{M+1} : \|S\|_2 = 1\}$$

We may assume that  $S^{(n)} \in \mathcal{S}$  since  $S_0^{(n)}$  is the price of time-0 consumption and should be greater than zero (otherwise, it will lead to  $V^L(\theta^L(S)) = -\infty$  as competitive traders demand infinite amount of time-0 consumption for free).  $\mathcal{S}$  is closed and bounded. Therefore, it is compact, which implies any sequence has a converging subsequence. There is a subsequence of  $S^{(1)}, S^{(2)}, \dots, S^{(n)} \dots$  in  $\mathcal{S}$  that converges. Let this subsequence be  $P^{(1)}, P^{(2)}, \dots$ , which converges to the limit  $\hat{S} \in \mathcal{S}$ .

Proposition 1 suggests that  $\theta(S)$ ,  $\theta^L(S)$  and  $c(S)$  are well-defined smooth functions on  $\mathbb{R}^{M+1} \setminus \mathcal{A}$ . If  $\hat{S} \notin \mathcal{A}$ , by continuity,

$$\lim_{n \rightarrow \infty} \theta^L(P^{(n)}) = \theta^L(\hat{S})$$

This suggests that  $V^L(\theta^L(\hat{S})) = \hat{V}^L$  and  $\hat{S}$  is the large trader equilibrium price.

We shall prove  $\hat{S} \notin \mathcal{A}$  by contradiction. Suppose  $\hat{S} \in \mathcal{A}$ . Firstly, we must have  $\|c^i(P^{(n)})\|_2 \rightarrow \infty$ . This is because if  $\|c^i(P^{(n)})\|_2 < M \forall n$  for some constant  $M$ ,  $c^i(P^{(n)})$  lies in a compact set  $B(M)$  where  $B(M)$  denotes a closed ball in  $\mathbb{R}^{K+1}$  with radius  $M$  in Euclidean norm. Then there is a converging subsequence  $c^i(P^{(n_1)}), c^i(P^{(n_2)}), \dots$  such that  $c^i(P^{(n_i)}) \rightarrow \hat{c}^i \in B(M)$ . Note that F.O.C for competitive trader  $i$ 's optimization problem gives

$$\partial U^i(c^i(P^{(n_i)})) X^+ = \lambda^{i(l)} P^{(n_i)} \tau$$

Taking limit, by continuity of  $\partial U^i$  and convergence of  $P^{(n_l)}$ , we have

$$\begin{aligned}\partial U^i(\hat{c}^i) X^+ &= \lim_{l \rightarrow \infty} \lambda^{i(l)} \lim_{l \rightarrow \infty} P^{(n_l)^\top} \\ &= \hat{\lambda}^i \hat{S}^\top\end{aligned}$$

Since  $\hat{c}^i$  is finite,  $\partial U^i(\hat{c}^i) \gg 0$ .  $\partial U^i(\hat{c}^i) X^+$ 's first entry is  $u_c^i(\hat{c}_0^i) \neq 0$ , which suggests  $\hat{\lambda}^i \neq 0$ .  $\partial U^i(\hat{c}^i) / \hat{\lambda}^i$  is a positive state price vector that prices all securities. This implies no arbitrage, which contradicts  $\hat{S} \in \mathcal{A}$ . Hence,  $\|c^i(P^{(n)})\|_2 \rightarrow \infty$ .

However, by non-negative consumption assumption, we know that  $c_k^i(P^{(n)}) \geq 0$  for  $k = 0, 1, \dots, K$ . If  $\|c^i(P^{(n)})\|_2 \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} \max_{k=0,1,\dots,K} c_k^i(P^{(n)}) = +\infty$$

which implies that

$$\lim_{n \rightarrow \infty} \min_{k=0,1,\dots,K} c_k^L(P^{(n)}) = -\infty$$

This suggests that, for large  $n$ ,  $V^L(\theta^L(P^{(n)})) < V^L(\theta^L(\bar{S}))$ , which is again a contradiction. Hence,  $\hat{S}$  cannot be a price vector that allows arbitrage.

### Proof of Proposition 3:

$L$  solves  $\max_S V^L(-\theta(S))$ . F.O.C suggests that the large trader equilibrium price vector  $\hat{S}$  satisfy  $-\partial V^L(\theta(\hat{S})) \partial \theta(\hat{S}) = 0$ . By Walras law, one can show  $\hat{S}^\top \partial \theta(\hat{S}) = -\theta(\hat{S})$  (see Mas-Collel, Whinston and Green (1995)). As long as  $\theta \neq 0$ ,  $\partial V^L \neq \lambda \hat{S}^\top$  for some  $\lambda$ . This implies inefficiency (See Magill and Quinzii (1996)). Conversely, if  $\theta(\hat{S}) = 0$ , it implies that the optimal portfolio is still  $\theta^L = 0$  if large trader were to behave competitively. Hence, the allocation in the large trader equilibrium is the same as that in the competitive equilibrium with price-taking  $L$ , which is efficient.

### Proof of Proposition 4:

Consider the competitive equilibrium with price-taking  $L$ , in which the large trader were to behave competitively. She optimizes her security demand  $\theta^L$  under budget constraint  $\tilde{S}^\top \theta^L \leq 0$  where  $\tilde{S}$  is the equilibrium price vector. If she trades by submitting security demand  $\tilde{\theta}^L \neq 0$ , it implies that  $V^L(\tilde{\theta}^L) > V^L(0)$  since no trade is always a feasible choice for her. Returning to the large trader equilibrium, she will still trade ( $\hat{\theta}^L \neq 0$ ) since no trade is dominated by  $\tilde{\theta}^L$ , which can be implemented by setting the price to be  $\tilde{S}$ .

### Proof of Proposition 5:

Suppose all competitive traders have lower welfare under the large trader equilibrium price  $\hat{S}$  set by  $L$ . Then we must have

$$\hat{S}^\top \tilde{\theta}^i > 0 \text{ (Previous equilibrium portfolio no longer attainable)}$$

where  $\tilde{\theta}^i$  is the equilibrium security demand of competitive trader  $i$  in the competitive equilibrium without  $L$ . Summing over  $i$ , we have

$$\sum_{i=1}^N \hat{S}^\top \tilde{\theta}^i > 0$$

But  $\sum_{i=1}^N \tilde{\theta}^i = 0$  (competitive equilibrium requires market-clearing). So

$$\sum_{i=1}^N \hat{S} \tilde{\theta}^i = \hat{S}^\top \sum_{i=1}^N \tilde{\theta}^i = 0$$

Contradiction.

### Proof of Proposition 6:

It has been proven that, generically, the rank of  $\partial \theta(\bar{S})$  is  $M$  (see Mas-Collel, Whinston and Green (1995)). This suggests that, for any vector  $v \in \mathbb{R}^{M+1}$  such that  $v \neq k\bar{S}$  for some  $k \neq 0$ ,  $v^\top \partial \theta(\bar{S}) \neq 0$  (i.e. the only direction that will lead to  $v^\top \partial \theta(\bar{S}) = 0$  is the direction of equilibrium prices, which has to be the case by Walras' Law). If  $\partial U^L(e^L) X^+ \neq \lambda \bar{S}^\top$ , this implies that  $\partial U^L(e^L) X^+ \partial \theta(\bar{S}) \neq 0$ , which implies the F.O.C of the large trader's optimization is not satisfied. Hence, there exists some direction of perturbation on security price vector  $\bar{S}$  that leads to improvement of the large trader's welfare.

### Proof of Proposition 7:

Let us denote the second derivative of the indirect utility function of the large trader with respect to her security holding  $\theta^L$  as  $\partial^2 V(\theta^L)$ . By definition,

$$\partial^2 V^L(\theta^L) = (X^+)^\top \partial^2 U^L(e^L + X^+ \theta^L) X^+ \quad (2.24)$$

Note that

$$\partial^2 U^L = \begin{bmatrix} u_{cc}^L(c_0^L) & \dots & \dots & 0 \\ 0 & u_{cc}^L(c_1^L) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & u_{cc}^L(c_0^L) \end{bmatrix} \quad (2.25)$$

is negative-definite. Hence,  $V^L(\cdot)$  is also negative-definite, which implies that it is concave. Thus,

$$\begin{aligned} V^L(\theta^L(S)) - V^L(\theta^L(\bar{S})) &< \partial V^L(\theta^L(\bar{S})) (\theta^L(S) - \theta^L(\bar{S})) \\ &= \partial V^L(0) \theta^L(S) = \partial U^L(e^L) X^+ \theta^L(S) \\ &= \lambda \bar{S}^\top \theta^L(S) = -\lambda \bar{S}^\top \theta(S) \\ &\leq 0 \end{aligned} \quad (2.26)$$

by WARP at  $\bar{S}$ . Therefore, by setting price to be different from the prevailing one, the large trader achieves a lower utility. Thus, she will not manipulate.

Conversely, if WARP fails to hold at  $\bar{S}$ , there exists a price vector  $S$  such that  $\bar{S}^\top \theta(S) < 0$  ( $S^\top \theta(\bar{S}) = 0$  is automatically satisfied as  $\theta(\bar{S}) = 0$ ). By no trading motive, we have

$$\partial V^L(0) = \partial U^L(e^L) X^+ = \lambda \bar{S}^\top \quad (2.27)$$

Thus,  $\bar{S}^\top \theta(S) < 0$  implies  $\partial V^L(0) \theta^L(S) > 0$ . Consider

$$V^L(\theta^L(S)) - V^L(0) = \int_0^1 \partial V^L(\theta^L(S)t) \theta^L(S) dt \quad (2.28)$$

If the large trader is sufficiently risk-tolerant at  $e^L$ ,  $\|\partial^2 U^L(e^L)\|$  is small, which implies that  $\|\partial^2 V^L(0)\|$  is small. So  $\partial V^L(\theta^L(S)t) \theta^L(S) > 0$ , which implies

$$V^L(\theta^L(S)) > V^L(0)$$

This suggests that the large trader could manipulate price to  $S$  and improve her utility.

### Proof of Proposition 8:

A similar proof is given for goods market in Mas-Colell (1985). Walras' Law gives that

$$\forall S, S^\top \theta(S) = 0$$

So we have

$$(\bar{S} + \eta)^\top \theta(\bar{S} + \eta) = 0$$

Therefore,

$$\bar{S}^\top \theta(\bar{S} + \eta) = -\eta^\top \theta(\bar{S} + \eta) \quad (2.29)$$

By Fundamental Theorem of Calculus, we have

$$\begin{aligned} \theta(\bar{S} + \eta) &= \theta(\bar{S}) + \int_0^1 \partial \theta(\bar{S} + \eta t) \eta dt \\ &= \int_0^1 \partial \theta(\bar{S} + \eta t) \eta dt \end{aligned} \quad (2.30)$$

This suggests that

$$\begin{aligned} \bar{S}^\top \theta(\bar{S} + \eta) &= -\eta^\top \theta(\bar{S} + \eta) \\ &= -\int_0^1 \eta^\top \partial \theta(\bar{S} + \eta t) \eta dt \end{aligned} \quad (2.31)$$

Making use of the condition that  $\partial \theta(\bar{S})$  is not negative semi-definite, there exists a  $v \in \mathbb{R}^{M+1}$  such that  $\|v\|_2 = 1$  and  $v^\top \partial \theta(\bar{S}) v > 0$ . Since  $\partial \theta(S)$  is continuous by smoothness of  $\theta(S)$ , there exists an open ball  $B(\bar{S}, r)$  around  $\bar{S}$  with radius  $r > 0$  such that

$$v^\top \partial \theta(S) v > 0 \forall S \in B(\bar{S}, r)$$

Choose  $\eta = \min\left(\frac{\varepsilon}{2}, r\right) v$ . We have

$$\eta^\top \partial \theta (\bar{S} + \eta t) \eta > 0 \forall t \in [0, 1]$$

Plugging back into the integral, we obtain

$$\bar{S}^\top \theta (\bar{S} + \eta) = - \int_0^1 \eta^\top \partial \theta (\bar{S} + \eta t) \eta dt < 0 \quad (2.32)$$

Therefore, by setting  $S = \bar{S} + \eta$ , WARP fails at  $\bar{S}$ :  $\bar{S}^\top \theta (S) < 0$  and  $\|S - \bar{S}\|_2 = \min\left(\frac{\varepsilon}{2}, r\right) < \varepsilon$ . By Proposition 7, a sufficiently risk-tolerant large trader can improve her utility through local manipulation of the security price vector.

### Proof of Proposition 9:

Since  $X^+$  and  $X^{+'}$  are complete markets security structures, for all corresponding security prices  $S$  and  $S'$  that do not lead to arbitrage, we have state prices  $\phi$  and  $\phi' \in \mathbb{R}_{++}^{K+1}$  such that

$$\phi^\top X^+ = S \text{ and } \phi'^\top X^{+'} = S'$$

From competitive traders' point of view, they optimize consumption allocation using state prices implied by the security price vector. As a result, their aggregate excess consumption is a function of the state prices. The large trader maximizes her utility by picking the optimal state prices  $\hat{\phi}$ , which can be implemented some security price vectors  $\hat{S} = \hat{\phi}^\top X^+$  and  $\hat{S}' = \hat{\phi}^\top X^{+'}$  under security structures  $X^+$  and  $X^{+'}$  respectively. Since  $\hat{\phi}$  leads to one particular consumption allocation, the exact security structures do not matter.

### Proof of Proposition 10.1:

See Brennan & Kraus (1978).

### Proof of Proposition 10.2:

Monotonicity implies WARP is straightforward: suppose  $\phi^\top z(\phi') \leq 0$  and  $z(\phi) \neq z(\phi')$ , we need to show  $\phi'^\top z(\phi) > 0$  to ensure WARP. From monotonicity, we have

$$[z(\phi) - z(\phi')]^\top (\phi - \phi') < 0 \quad (2.33)$$

(strict inequality since  $z(\phi) \neq z(\phi')$ ). But

$$\begin{aligned} [z(\phi) - z(\phi')]^\top (\phi - \phi') &= z(\phi)^\top \phi - z(\phi')^\top \phi - z(\phi)^\top \phi' + z(\phi')^\top \phi' \\ &= -z(\phi')^\top \phi - z(\phi)^\top \phi' \text{ (since } z(\phi)^\top \phi = z(\phi')^\top \phi' = 0) \end{aligned} \quad (2.34)$$

Thus,  $[z(\phi) - z(\phi')]^\top (\phi - \phi') < 0$  implies  $z(\phi)^\top \phi' > -z(\phi')^\top \phi \geq 0$ . WARP is satisfied.

To obtain monotonicity condition, we need individual excess consumption  $z^i(\phi)$  to satisfy monotonicity. Then, summing over all the individuals, we have aggregate excess consumption  $z(\phi)$  also satisfy monotonicity. A sufficient condition for monotonicity of

individual demand with constant wealth is  $\max_{c \in R_{++}^{K+1}} \gamma^i(c) - \min_{c \in R_{++}^{K+1}} \gamma^i(c) < 4$ . (See Quah (2003) for a proof). If endowments are collinear, we can scale the state prices by the same constant for all individuals to keep their wealth levels unchanged under different state price vectors, which allows us to apply the above result with an endowment economy.

**Proof of Proposition 10.3:**

Arrow, Block and Hurwicz (1959) shows that global gross substitution implies WARP at  $\bar{\phi}$ . Hens and Loeffler (1995) shows  $\gamma^i(c) < 1 \forall c$  implies gross substitution.

**Proof of Proposition 10.4:**

If there is no trading in the competitive equilibrium without  $L$ , each competitive trader's excess consumption  $z^i(\bar{\phi}) = 0$ . Since WARP holds at the level of individual trader,  $\phi'^T z^i(\bar{\phi}) = \phi'^T 0 = 0$  implies that  $\bar{\phi}^T z^i(\phi') > 0$  if  $z^i(\phi') \neq z^i(\bar{\phi})$ . Summing over  $i$ , we have

$$\bar{\phi}^T z(\phi') = \bar{\phi}^T \sum_{i=1}^N z^i(\phi') = \sum_{i=1}^N \bar{\phi}^T z^i(\phi') > 0$$

WARP holds at  $\bar{\phi}$ .

**Proof of Proposition 10.5:**

Suppose  $L$  sets a different state price vector  $\phi \neq \bar{\phi}$ . For each competitive trader  $i$ , there exists a consumption bundle  $\xi^i$  with  $\xi^i = k^i 1$  such that  $\phi^T \xi^i = \phi^T e^i$  (i.e.  $\xi^i$  is a riskless consumption bundle attainable under state prices  $\phi$ ). Expected utility preference implies that  $\partial U^i(\xi^i) = u_c^i(k^i) (1 \ p_1 \dots p_K)^T$  and  $\partial U^i(\bar{c}^i) = u_c^i(\bar{c}_0^i) (1 \ p_1 \dots p_K)^T$ . Thus,  $\partial U^i(\xi^i)$  is proportional to  $\partial U^i(\bar{c}^i)$ . But  $\partial U^i(\bar{c}^i)$  is proportional to  $\bar{\phi}$  in complete markets.  $\partial U^i(\xi^i)$  is proportional to  $\bar{\phi}$  as well. Since  $\phi \neq \bar{\phi}$ ,  $\partial U^i(\xi^i)$  is not proportional to  $\phi$ , which implies  $\xi^i$  is not agent  $i$ 's optimal consumption under state price vector  $\bar{\phi}$ . So the optimal consumption  $c^i(\phi)$  delivers higher utility than  $\xi^i$ :  $U(c^i(\phi)) > U(\xi^i)$ . By supporting hyperplane theorem, we have  $\partial U^i(\xi^i) c^i(\phi) > \partial U^i(\xi^i) \xi^i$ .

Since  $\partial U^i(\xi^i) = \frac{u_c^i(k^i)}{u_c^i(\bar{c}_0^i)} \partial U^i(\bar{c}^i) = \lambda^i \bar{\phi}$  for some  $\lambda^i > 0$ ,

$$\bar{\phi}^T c(\phi) = \bar{\phi}^T \sum_{i=1}^N c^i(\phi) > \bar{\phi}^T \sum_{i=1}^N \xi^i \quad (2.35)$$

Next, we want to show  $\sum_{i=1}^N \xi^i = \sum_{i=1}^N e^i$ . Since  $\phi^T \xi^i = \phi^T e^i$ ,

$$\phi^T \left[ \left( \sum_{i=1}^N \xi^i \right) - \left( \sum_{i=1}^N e^i \right) \right] = 0$$

But  $\left(\sum_{i=1}^N \xi^i\right) = \left(\sum_{i=1}^N k^i\right) 1 = k_\xi 1$  and  $\left(\sum_{i=1}^N e^i\right) = \left(\sum_{i=1}^N \bar{c}^i\right) = k_e 1$  (complete risk-sharing without  $L$ ), so

$$\phi^\top \left[ \left( \sum_{i=1}^N \xi^i \right) - \left( \sum_{i=1}^N e^i \right) \right] = (k_\xi - k_e) (\phi^\top 1) = 0 \quad (2.36)$$

This implies  $k_\xi = k_e$  since  $\phi^\top 1 > 0$ . So we have  $\sum_{i=1}^N \xi^i = \sum_{i=1}^N e^i = e$ . As a result,

$$\begin{aligned} \bar{\phi}^\top c(\phi) &> \bar{\phi}^\top \sum_{i=1}^N \xi^i = \bar{\phi}^\top e \\ \bar{\phi}^\top z(\phi) &= \bar{\phi}^\top (c(\phi) - e) > 0 \end{aligned} \quad (2.37)$$

which suggests that WARP holds at  $\bar{\phi}$ .

The equivalence between complete risk-sharing and no aggregate risk is known for a long time. See Malinvaud (1972) for a discussion.

### Proof of Proposition 11:

F.O.C of competitive trader  $i$ 's utility optimization problem gives:

$$u_c^i(c_k^i) = \lambda^i \pi_k = u_c^i(c_0^i) \frac{\pi_k}{\pi_0} \quad (2.38)$$

Note that here state-0 is used on the right-hand side. However, this is not necessary. We can use any other state and obtain the same result.

Differentiating w.r.t  $\phi_j$  and re-arranging:

$$\frac{\partial c_k^i}{\partial \phi_j} = \frac{\alpha^i(c_0^i)}{\alpha^i(c_k^i)} \frac{\partial c_0^i}{\partial \phi_j} \text{ for } j \neq 0, k \quad (2.39)$$

$$\frac{\partial c_k^i}{\partial \phi_0} = \frac{\alpha^i(c_0^i)}{\alpha^i(c_k^i)} \frac{\partial c_0^i}{\partial \phi_0} + \frac{1}{\alpha^i(c_k^i)} \frac{1}{\phi_0} \text{ for } j = 0 \quad (2.40)$$

$$\frac{\partial c_k^i}{\partial \phi_k} = \frac{\alpha^i(c_0^i)}{\alpha^i(c_k^i)} \frac{\partial c_0^i}{\partial \phi_k} - \frac{1}{\alpha^i(c_k^i)} \frac{1}{\phi_k} \text{ for } j = k \quad (2.41)$$

Competitive trader's budget constraint is always tight given non-satiation. So we have

$$\sum_{k=0}^K \phi_k c_k = \sum_{k=0}^K \phi_k e_k \quad (2.42)$$

Differentiating the budget constraint w.r.t.  $\phi_j$ :

$$\sum_{k=0}^K \phi_k \frac{\partial c_k^i}{\partial \phi_j} = e_j^i - c_j^i = -z_j^i$$

Substituting  $\frac{\partial c_k^i}{\partial \phi_j}$  into the above equation, we get

$$\frac{\partial c_k^i}{\partial \phi_j} = \frac{\frac{1}{\alpha^i(c_j^i)} - z_j^i}{\alpha^i(c_k^i) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha^i(c_l^i)} \right)} \text{ for } j \neq k \quad (2.43)$$

$$\frac{\partial c_k^i}{\partial \phi_k} = \frac{\frac{1}{\alpha^i(c_k^i)} - z_k^i}{\alpha^i(c_k^i) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha^i(c_l^i)} \right)} - \frac{1}{\alpha^i(c_k^i)} \frac{1}{\phi_k} \text{ otherwise} \quad (2.44)$$

**Proof of Proposition 12:**

With unit wealth, trader  $i$ 's consumption demand

$$h^i(\phi) = \arg \max_c \sum_{k=0}^K p_k \frac{c_k^{1-\gamma^i}}{1-\gamma^i} \text{ s.t. } \phi^\top c \leq 1 \quad (2.45)$$

F.O.C gives

$$(h^i)^{-\gamma^i} = \lambda^i \pi$$

which yields

$$h^i = (\lambda^i)^{-\frac{1}{\gamma^i}} \pi^{-\frac{1}{\gamma^i}} \quad (2.46)$$

Substitute into the budget constraint, which binds by non-satiation:

$$\phi^\top h^i = (\lambda^i)^{-\frac{1}{\gamma^i}} \left( \phi^\top \pi^{-\frac{1}{\gamma^i}} \right) = 1$$

which gives

$$(\lambda^i)^{-\frac{1}{\gamma^i}} = \frac{1}{\left( \phi^\top \pi^{-\frac{1}{\gamma^i}} \right)} \quad (2.47)$$

Hence,

$$h^i(\phi) = \frac{\pi^{-\frac{1}{\gamma^i}}}{\left( \phi^\top \pi^{-\frac{1}{\gamma^i}} \right)} \quad (2.48)$$

To show the decomposition, note that  $c^i(\phi) = \zeta^i(\phi, \phi^\top e^i)$ . Thus, we have  $\partial z^i = \partial c^i = \partial \zeta^i$ . But  $\zeta^i(\phi, \phi^\top e^i) = (\phi^\top e^i) h^i(\phi)$ . So

$$\partial z^i = (\phi^\top e^i) \partial h^i + h^i e^{i\top} \quad (2.49)$$



Upon differentiation, we have

$$\begin{aligned}\frac{\partial h_k^i}{\partial \phi_j} &= -\left(1 - \frac{1}{\gamma^i}\right) \frac{\phi_j^{-\frac{1}{\gamma^i}} p_j^{\frac{1}{\gamma^i}} \pi_k^{-\frac{1}{\gamma^i}}}{\left(\phi^\top \pi^{-\frac{1}{\gamma^i}}\right)^2} \text{ if } j \neq k \\ &= -\left(1 - \frac{1}{\gamma^i}\right) h_j^i h_k^i\end{aligned}\quad (2.50)$$

and

$$\begin{aligned}\frac{\partial h_k^i}{\partial \phi_k} &= -\left(1 - \frac{1}{\gamma^i}\right) \frac{\phi_k^{-\frac{1}{\gamma^i}} p_k^{\frac{1}{\gamma^i}} \pi_k^{-\frac{1}{\gamma^i}}}{\left(\phi^\top \pi^{-\frac{1}{\gamma^i}}\right)^2} - \frac{1}{\gamma^i} \frac{\phi_k^{-\frac{1}{\gamma^i}-1} p_k^{\frac{1}{\gamma^i}}}{\left(\phi^\top \pi^{-\frac{1}{\gamma^i}}\right)} \text{ if } j = k \\ &= -\left(1 - \frac{1}{\gamma^i}\right) h_k^i h_k^i - \frac{1}{\gamma^i} h_k^i / \phi_k\end{aligned}\quad (2.51)$$

In matrix form, we have

$$\partial h^i = -\left(1 - \frac{1}{\gamma^i}\right) h^i h^{i\top} - \frac{1}{\gamma^i} \text{diag}(h^i / \phi) \quad (2.52)$$

where  $h^i / \phi$  is a vector formed by coordinate-wise division and  $\text{diag}(v)$  transforms a vector into a square matrix with diagonal being  $v$  and 0 elsewhere. Substituting into (2.48), we have

$$\begin{aligned}\partial z^i &= w^i \left[ -\left(1 - \frac{1}{\gamma^i}\right) h^i h^{i\top} - \frac{1}{\gamma^i} \text{diag}(h^i / \phi) \right] + h^i e^{i\top} \\ &= w^i \left[ -\left(1 - \frac{1}{\gamma^i}\right) h^i h^{i\top} - \frac{1}{\gamma^i} \text{diag}(h^i / \phi) + h^i h^{i\top} \right] - w^i h^i h^{i\top} + h^i e^{i\top} \\ &= w^i \left[ -\frac{1}{\gamma^i} (\text{diag}(h^i / \phi) - h^i h^{i\top}) \right] - h^i (c^i - e^i)^\top \\ &= w^i s^i(\phi) - h^i z^{i\top}\end{aligned}\quad (2.53)$$

### Proof of Proposition 13:

For competitive investors, F.O.C is still

$$u_c(c_k) = \lambda \pi_k = u_c(c_0) \frac{\pi_k}{\pi_0}$$

Differentiating w.r.t  $\phi_j$ , we have a new term that is engendered by the dependence of

probability distribution on state price  $\phi$

$$\frac{\partial c_k}{\partial \phi_j} = \frac{\alpha(c_0)}{\alpha(c_k)} \frac{\partial c_0}{\partial \phi_j} + \frac{1}{\alpha(c_k) \tilde{p}_k} \frac{\partial \tilde{p}_k}{\partial \phi_j} \quad (2.54)$$

$$\frac{\partial c_k}{\partial \phi_k} = \frac{\alpha(c_0)}{\alpha(c_k)} \frac{\partial c_0}{\partial \phi_k} + \frac{1}{\alpha(c_k) \tilde{p}_k} \frac{\partial \tilde{p}_k}{\partial \phi_j} - \frac{1}{\alpha(c_k)} \frac{1}{\phi_k} \quad (2.55)$$

Differentiating the budget constraint w.r.t.  $\phi_j$ , we still have

$$\sum_{k=0}^K \phi_k \frac{\partial c_k}{\partial \phi_j} = e_j - c_j = -z_j$$

In equilibrium,  $c = e$ ,  $z = 0$  and  $\tilde{p} = p$ . This gives

$$\begin{aligned} \frac{\partial c_k(\bar{\phi})}{\partial \phi_j} &= \frac{\frac{1}{\alpha(e_j)} - \sum_{l=0}^K \frac{\pi_l}{\alpha(e_l)} \frac{\partial \tilde{p}_l}{\partial \phi_j}}{\alpha(e_k) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha(e_l)} \right)} + \frac{1}{\alpha(e_k) p_k} \frac{\partial \tilde{p}_k}{\partial \phi_j}, j \neq k \\ &= \frac{\frac{1}{\alpha(e_j)} + \sum_{l=0, l \neq k}^K \frac{\phi_l}{\alpha(e_l)} \frac{\partial \ln(\tilde{p}_k/\tilde{p}_l)}{\partial \phi_j}}{\alpha(e_k) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha(e_l)} \right)} \end{aligned} \quad (2.56)$$

$$\begin{aligned} \frac{\partial c_k}{\partial \phi_k} &= \frac{\frac{1}{\alpha(e_k)} - \sum_{l=0}^K \frac{\pi_l}{\alpha(e_l)} \frac{\partial \tilde{p}_l}{\partial \phi_k}}{\alpha(e_k) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha(e_l)} \right)} + \frac{1}{\alpha(e_k) \tilde{p}_k} \frac{\partial \tilde{p}_k}{\partial \phi_j} - \frac{1}{\alpha(e_k)} \frac{1}{\phi_k} \\ &= \frac{\sum_{l=0, l \neq k}^K \frac{\phi_l}{\alpha(e_l)} \left( \frac{\partial \ln(\tilde{p}_k/\tilde{p}_l)}{\partial \phi_k} - \frac{1}{\phi_k} \right)}{\alpha(e_k) \left( \sum_{l=0}^K \frac{\phi_l}{\alpha(e_l)} \right)} \end{aligned} \quad (2.57)$$

**Proof of Proposition 16:**

Define  $\Delta := S - \bar{S}$  and expand  $\partial U^L(e^L)^\top X(S) \theta(S)$  around  $\bar{S}$ :

$$\begin{aligned}
& \partial U^L (e^L)^\top X(S) \theta(S) \\
= & \partial U^L (e^L)^\top [X(\bar{S}) \theta(\bar{S}) + \underbrace{\left( \sum_{m \in \mathbb{M}(\xi), \xi \in \mathbb{D}(\xi_0)} \Delta_m(\xi) \frac{\partial X}{\partial S_m(\xi)} \right)}_{\text{First Order 1}} \theta(\bar{S}) + \quad (2.58) \\
& + \underbrace{X(\bar{S}) \partial \theta \Delta}_{\text{First Order 2}} + \underbrace{X(\bar{S}) \left( \sum_{m \in \mathbb{M}(\xi), \xi \in \mathbb{D}(\xi_0)} \Delta_m(\xi) \frac{\partial [\partial \theta]}{\partial S_m(\xi)} \right) \Delta}_{\text{Second Order 1}} + \\
& + \underbrace{\left( \sum_{m \in \mathbb{M}(\xi), \xi \in \mathbb{D}(\xi_0)} \sum_{m' \in \mathbb{M}(\xi'), \xi' \in \mathbb{D}(\xi_0)} \Delta_m(\xi) \Delta_{m'}(\xi') \frac{\partial^2 X}{\partial S_m(\xi) \partial S_{m'}(\xi')} \right) \theta(\bar{S})}_{\text{Second Order 2}} \\
& + \underbrace{2 \left( \sum_{m \in \mathbb{M}(\xi), \xi \in \mathbb{D}(\xi_0)} \Delta_m(\xi) \frac{\partial X}{\partial S_m(\xi)} \right) \partial \theta \Delta + \dots}_{\text{Second Order 3}} \quad (2.59)
\end{aligned}$$

As  $\partial U^L (e^L)^\top X(\hat{S}) = 0$  (no trading motive), the first order effect 2 and second order effect 1 are all 0. Since  $\bar{S}$  is a market-clearing equilibrium price,  $\theta(\bar{S}) = 0$ . So first order effect 1 and second order effect 2 are all 0.

For each security  $m$  tradable at  $\xi$ , there are 2 cases:

1. It is the first time that  $m$  gets traded
2. There is a preceding node  $\xi'$ , at which  $m$  gets traded directly before  $\xi$  (then  $S_m(\xi)$  determines the payoff of holding 1 unit of security  $m$  from  $\xi'$  to  $\xi$ )

Hence,

$$\frac{\partial X}{\partial S_m(\xi)} = \xi \rightarrow \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & -1 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ & & & (m, \xi) & \end{bmatrix} \text{ in case 1} \quad (2.60)$$

$$= \xi \rightarrow \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & -1 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ & (m, \xi') & & (m, \xi) & \end{bmatrix} \text{ in case 2} \quad (2.61)$$

there is an '1' in case 2 because  $S_m(\omega)$  also determines the payoff (including capital gain) of holding 1 unit of  $m$  from the last trading period  $\xi'$ .

Hence,

$$\begin{aligned}
\frac{1}{2} \partial U^L (e^L)^\top X(S) \theta(S) &\approx \partial U^L (e^L)^\top \left( \sum_{m \in \mathbb{M}(\xi), \xi \in \mathbb{D}(\xi_0)} \Delta_m(\xi) \frac{\partial X}{\partial S_m(\xi)} \right) \partial \theta \Delta \\
&= \left( \sum_{m \in \mathcal{M}(\omega), \omega} \Delta_m(\xi) \partial U^L (e^L)^\top \frac{\partial X}{\partial S_m(\xi)} \right) \partial \theta \Delta \\
&= \Delta^\top (B - I) D^L \partial \theta \Delta
\end{aligned} \tag{2.62}$$

Therefore, if  $(I - B) D^L \partial \theta$  is NOT negative semi-definite, can find a direction  $\Delta$  such that moving the security price vector  $S$  along this direction will lead to  $\partial U^L (e^L)^\top X(S) \theta(S) < 0$ . By Proposition 15, this gives rise to pure manipulation.

## References

- Aggarwal, R.K. and G. Wu 2006, "Stock market manipulations," *Journal of Business*, 79, 1915-1953
- Allen, F. and D. Gale 1992, "Stock-price Manipulation," *Review of Financial Studies*, 5, 503-529
- Allen, F., L. Litov, L. and J. Mei 2006, "Large investors, price manipulation, and limits to arbitrage: An anatomy of market corners," *Review of Finance*, 10, 645-693
- Arrow, K., H. Block and L. Hurwicz, 1959, "On the stability of the competitive equilibrium II," *Econometrica*, 27, 82-109
- Arrow, K.J. and G. Gerard 1954, "Existence of an equilibrium for a competitive economy," *Econometrica*, 22, 265-290
- Basak, S. 1997, "Consumption choice and asset pricing with a non-price-taking agent," *Economic Theory*, 10, 437-462
- Bagnoli, M. and B. Lipman 1996, "Stock price manipulation through take-over bids," *RAND Journal of Economics*, 27, 124-147
- Benabou, R. and G. Laroque, 1992, "Using privileged information to manipulate markets: insiders, gurus, and creditability," *Quarterly Journal of Economics*, 107, 921-958
- Bonanno, G. 1994, "General equilibrium theory with imperfect competition," *Journal of Economic Surveys*, 4, 297-328
- Brennan, M. and A. Kraus 1978, "Necessary conditions for aggregation in security markets," *Journal of Financial and Quantitative Analysis*, 13, 407-418
- Chan, L. and L. Kogan 2002, "Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices," *Journal of Political Economy*, 110, 1255-1285
- Diamond, P.A. 1967, "The role of a stock market in a general equilibrium model with technological uncertainty," *American Economic Review*, 57, 759-776
- Duffie, D. and C.F. Huang 1985, "Implementing Arrow-Debreu equilibria by continuous trading of a few long-lived securities," *Econometrica*, 53, 1337-1356
- Eom, K.S., E.J. Lee, and K.S. Park 2009, "Microstructure-based manipulation: Strategic behavior and performance of spoofing traders," *Working Paper*
- Fishman, M. and K. Hagerty 1995, "The mandatory disclosure of trades and market liquidity," *Review of Financial Studies*, 18, 417-457
- Giraud, G. and H. Stahn 2003, "Efficiency and imperfect competition with incomplete markets," *Journal of Mathematical Economics*, 39, 559-583

- Goldstein, I. and A. Guebel 2008, "Manipulation and the allocational role of prices," *Review of Economic Studies*, 75, 133-164
- Hart, O. 1977, "On the profitability of speculation," *Quarterly Journal of Economics*, 91, 579-597
- Hart, O. 1985, "Imperfect competition in general equilibrium: an overview of recent work," *Frontiers of Economics*
- Hens, T. and A. Loeffler 1995, "Gross substitution in financial markets," *Economics Letters*, 49, 34-43
- Hens, T. and B. Pilgrim 2010, "General equilibrium foundations of finance: Structure of incomplete markets models," *Springer*
- Jarrow, R. 1992, "Market manipulation, bubbles, corners, and short squeezes," *Journal of Financial and Quantitative Analysis*, 27, 311-336
- Jarrow, R. 1994, "Derivative security markets, market manipulation, and option pricing theory," *Journal of Financial and Quantitative Analysis*, 29, 241-261
- John, K. and R. Narayanan 1997, "Market manipulation and the role of insider trading regulations," *Journal of Business*, 70, 217-247
- Kumar, P. and D. Seppi 1992, "Futures manipulation with "cash settlement"," *Journal of Finance*, 47, 1485-1502
- Khwaja, A.I. and A. Mian 2005, "Unchecked intermediaries: Price manipulation in an emerging stock market," *Journal of Financial Economics*, 78, 203-241
- Kyle, A. 1985, "Continuous auction and insider trader," *Econometrica*, 53, 1315-1335
- Kyle, A. 1984, "A theory of futures market manipulations," *The Industrial Organization of Futures Markets*
- Kyle, A. and S. Viswanathan 2008, "How to define illegal manipulation," *American Economic Review: Papers and Proceedings*, 98, 274-279
- Magill, M. and M. Quinzii 1996, "Theory of incomplete markets," *The MIT Press*
- Mas-Collel, A. 1985, "The theory of general economic equilibrium: A differential approach," *Cambridge University Press*
- Mas-Collel, A., M. Whinston and J. Green 1995, "Microeconomics," *Oxford University Press*
- Maulinvaud, 1972, "The allocation of individual risks in large markets," *Journal of Economic Theory*, 4, 312-328

- Mei, J.P., G.J Wu and C.S Zhou 2004, "Behavior based manipulation: Theory and prosecution evidence," *Working Paper*
- Quah, J. 2003, "The law of demand and risk aversion," *Econometrica*, 73, 713-721
- Radner, R. 1978, "Rational expectation equilibrium: Generic existence and the information revealed by prices," *Econometrica*, 47, 655-678
- Shleifer, A. 1986, "Do demand curves for stocks slope down?," *Journal of Finance*, 41, 579-590
- Stahn, H. 1998, "On monopolistic equilibria with incomplete markets: the case of an exchange economy," *Journal of Mathematical Economics*, 29, 83-107
- Van Bommel, J. 2003, "Rumors," *Journal of Finance*, 58, 1499-1519
- Vayanos, D. 2001, "Strategic Trading in a dynamic noisy market," *Journal of Finance*, 56, 131-171
- Vila, J. 1989, "Simple games of market manipulation," *Economics Letters*, 29, 21-26
- Wang, J. 1996, "The term structure of interest rates in a pure-exchange economy with heterogeneous investors," *Journal of Financial Economics*, 41, 75-110

## Chapter 3

# Dynamic Trading With Price Impact In The Presence of Momentum Traders

(This is joint work with Jiang Wang)

### 3.1 Introduction

Large institutional investors often need to trade in markets that offer only limited liquidity. For instance, an U.S. domestic equity fund may trade small-cap stocks to load on size premium. Alternatively, an international equity fund may seek good investment opportunities or diversification from emerging market stocks. When they trade in such markets with small turnover, their trading may generate considerable price impact. Various studies have been performed on optimal trading with price impact in general or partial equilibrium frameworks before<sup>1</sup>.

This paper adds an unique element to this problem by noticing that such small markets are often populated with investors following momentum strategies due to potential existence of private information. Under certain circumstances, momentum strategies may be well-justified. For example, Hong, Lim and Stein (2000) finds that momentum strategies generate highest profit with small-cap stocks because of gradual information diffusion. However, such momentum strategies are often applied without well-founded reasons. It might be blind implementation of "technical analysis" or, in general, oversimplistic extrapolation of past return patterns. Alternatively, momentum sentiment might be simply out of investors' "irrational exuberance". Emerging markets, such as the Chinese stock market, have seen tremendous holding demand from individual investors after periods of rapid rise in stock price. Anecdotal evidence suggests many investors bought stocks simply after seeing their acquaintances make a fortune and hoped that the good returns would continue. In this paper, we study the existence of imperfectly rational momentum

---

<sup>1</sup>For e.g. Bertsimas and Lo (1998), Vayanos (2001)



traders, who suffer from "momentum bias" – their expectation of future return is higher than the true one when past returns have been higher than usual.

In a market with momentum traders, it is interesting to study how a large investor would trade optimally. Large investor has price impact. Price goes up when she buys and goes down when she sells. Since momentum traders' expectation about future return and, consequently, their current demand depend on prices in the past, the dynamic trading behavior by a large investor with price impact is fundamentally different from when she trades against rational traders without momentum bias. This paper considers an equilibrium trading model in discrete time with infinite horizon. Shocks to dividend are i.i.d over time. Thus, past returns and prices cannot forecast future shocks. There is no information asymmetry. Investors are risk-averse. There are two broad classes of investors: a non-competitive large investor and competitive small investors. Competitive investors can be fully rational or imperfectly rational with momentum bias. Large investor and rational investors fully optimize dynamically. Imperfectly rational investors have a momentum bias in their expectation of future return that leads their trading strategy to deviate from being fully optimal. To ensure that our results are not generated by peculiar model assumptions, we keep everything endogenous except the form of momentum bias, which is assumed to be a linear function of past realized excess returns.

We found several interesting behaviors by the large investor in the presence of momentum investors. When trading towards a target holding position, a risk-averse large trader will split her trade into smaller ones. If she trades against rational competitive investors without momentum bias, the small orders are in the same direction and the order size is decreasing over time. However, this is no longer the case when she trades against momentum investors. The order size may not be monotonically decreasing. She may slow down trading when momentum demand is driven up by her previous trades and accelerate subsequently when momentum demand dies down. Depending on the strength of momentum demand, the direction of her trades may start to alternate. She may engage in round-trip trades with declining amplitude to profit from momentum demand before settling towards the target level. Also, under certain circumstances, the large investor engages in "reverse trading" that aims at manipulating the intertemporal demand curve. For instance, when she plans to sell a large amount, she may buy a moderate amount first to create high realized return. Momentum bias will push the demand curve up next period when she sells. Her initial purchase might end up with a loss. But this is more than compensated by the gain from demand curve shift when she sells a large amount later on. Thirdly, a large investor tries to conserve the momentum demand by taking profit of it less aggressively or even incurring a loss in the short-run so that the momentum cycles will last longer with greater magnitude. And she benefits from more sustainable manipulation of the momentum sentiment. In comparison, if we replace her with a group of competitive rational investors, they will jump on the mispricing created by the momentum demand and take profit aggressively. Momentum sentiment declines quickly. Finally, we find that the excess return process actually exhibits strong reversal in the presence of momentum investors. Thus, momentum trading cannot be self-fulfilling in a rational equilibrium framework with long-lived arbitrageurs.

In addition, since we adopt a general equilibrium framework, the price impact function is generated endogenously by demand of competitive investors. This contrasts with partial equilibrium approach, which specifies price impact exogenously<sup>2</sup>. The implications could be very different. Several studies employed an exogenous price impact function and found that when price volatility increases, the large investor's trading speed increases given she is risk-averse. In our model, we find that price volatility does not affect the trading speed because the endogenously generated price impact becomes larger due to competitive investors' risk aversion, which cancels out the risk aversion effect on the large investor.

How large investors trade with price impact have been examined extensively before. In an equilibrium framework, Kyle (1985) pioneers such studies with an information monopolist, who trades to minimize price impact. Vayanos (2001) studies how a risk-averse large investor trades against rational competitive investors to reduce risk exposure with asymmetric information about her holding position. Our setup is close to Vayanos (2001) and the "trade splitting" effect is indeed similar. However, in both Kyle and Vayanos, the investors, whom the large investor trades against, are fully rational. Our focus is on the interaction between large investor and momentum investors. And the resulting behavior of the large investor is very different. With exogenously given price impact function, Bertsimas and Lo (1998) finds that a risk-neutral large investor should optimally divide her trade into orders of equal size. Almgren and Chriss (2000), Stanzl and Huberman (2005) extend to a large investor with risk-aversion and find that the order size should decrease as the deviation from the desired holding level gets smaller, which is similar to our result when the large investor trades against fully rational competitive investors. However, as mentioned above, the price impact function in our study is generated endogenously rather than exogenously given. Overall, these studies feature a large investor with genuine informational or allocative trading motive, who engages in "benign" manipulation<sup>3</sup> in the sense that the large investor restricts quantities traded to seek better transaction prices.

More severe price manipulation schemes have also been proposed. Allen and Gale (1992) classifies them into "information-based", "action-based" and "trade-based". "Information-based" manipulation requires information asymmetry. Manipulator releases false signals to mislead investors (Vila (1989), Benabou and Laroque (1992), Fishman and Hagerty (1995) etc.). "Action-based" manipulation encompasses actions that change the actual or perceived value of the underlying firms. Bagnoli and Lipman (1996) studies a fake take-over bid and Goldstein and Guebel (2008) analyzes how price manipulation can lead firm managers to forgo positive NPV projects. "Trade-based" manipulation requires simply trading. Allen and Gale (1992) provides a classic example, in which an uninformed manipulator pretends to be informed and takes advantage of the uninformed competitive investors. Mei, Wu and Zhou (2004) shows that a large investor can exploit behavioral bias among competitive investors to generate positive manipulation profit. While they study the disposition effect, our model focuses on the momentum sentiment.

There has been previous literature discussing the effect of momentum or positive-feedback

---

<sup>2</sup>For e.g. Almgren and Chriss (2000), Stanzl and Huberman (2005)

<sup>3</sup>See Kyle and Viswanathan (2008).

trading. In particular, Delong et al (1990) studies the positive feedback traders in a 3-period model. Barberis and Shleifer (2003) studies the "fund-switchers" who tend to switch to funds that perform better in the past. In their studies, competitive rational investors trade against imperfectly rational investors with momentum sentiment. Our model emphasizes on the presence of a large investor, who has ability to move price.

The paper proceeds as follows. In section 2, we describe the model. In section 3, we define the notion of equilibrium and outline the solution to the equilibrium. In section 4, we analyze the equilibrium results. Section 5 concludes.

## 3.2 Model Setup

We consider an economy in discrete-time setup with infinite horizon  $t = 0, 1, 2, \dots$ . There is a single consumption good used as the numeraire. There are two broad classes of investors we shall consider. Namely, they are competitive small investors (called S-Investors) and a non-competitive large investor (called L-Investor). S-Investors in general may have a bias in their expectation of future returns due to momentum or "trend-chasing" sentiment. When S-Investors have non-zero bias, they are imperfectly rational investors (called I-Investors). Otherwise, they are perfectly rational (called R-Investors). Behaviors of each type of investors will be explained later on.

### 3.2.1 Investment Opportunities

1. A riskless asset with constant interest rate  $r > 0$  per period. Denote the gross rate of return as  $R = 1 + r$ .
2. A tradable stock that pays a dividend  $D_t$  per period. The dividend follows an AR(1) process

$$D_{t+1} = a_D D_t + \varepsilon_{t+1} \text{ with } 0 \leq a_D \leq 1 \quad (3.1)$$

where  $\varepsilon_{t+1}$  are i.i.d idiosyncratic shocks. For simplicity, we assume that  $\varepsilon_t \sim N(0, \sigma_D^2)$ . Every investor has access to the stock market. The market for stock trading is cleared in a Walrasian manner. Let the total supply of the stock be  $\bar{\theta}$  and, without loss of generality, we can set  $\bar{\theta} = 1$ . Also, the ex-dividend price of stock at time  $t$  is  $P_t$ . Define the excess return of stock over the riskless security as

$$Q_t = P_t + D_t - R P_{t-1} \quad (3.2)$$

This is the extra gain of holding 1 share from period  $t-1$  to  $t$  relative to putting the same amount of wealth in riskless asset. Define the fundamental value of a share to be

$$F_t = E_t \left( \sum_{s=1}^{\infty} D_{t+s} R^{-s} \right) \quad (3.3)$$

, which is the present value of the future dividends discounted at the risk-free rate. It can be easily shown that

$$F_t = aD_t \text{ where } a = \frac{a_D}{R - a_D} \quad (3.4)$$

We can decompose the stock price  $P_t$  into  $P_t = \hat{P}_t + F_t$ , where  $\hat{P}_t$  is the risk discount of the stock price relative to the fundamental value.

### 3.2.2 Investors

All investors maximize time-additive constant absolute risk aversion (CARA) utility

$$E_t^i \left[ - \sum_{s=0}^{\infty} \beta^s e^{-\gamma^i c_{t+s}^i} \right] \text{ where } i = S, L$$

All of them have the same discount factor  $\beta$ . However, investor of type  $i$  has risk aversion  $\gamma^i$ . Define the effective risk aversion to be  $\alpha^i = \frac{\gamma^i}{R}$ .

The expectation operator for the rational R-Investors and the non-competitive rational L-Investor is the same as the true expectation operator  $E[\cdot]$ . We use  $E^S[\cdot]$  to denote the expectation operator for a general S-Investor, who may suffer from the momentum bias and have an expectation different from the true one.

Competitive investors (S-Investors) are identical. They are infinitesimally small and form a continuum indexed by  $i \in [0, \omega_S]$ , where  $\omega_S$  is the aggregate weight of small investors. Without loss of generality, we can set  $\omega_S = 1$ . Consequently, they are price takers, who submit demand schedules competitively without having to consider their own impact on price. Since they are exactly identical, we may treat them as a single investor who behaves competitively. When S-Investors are fully rational, they are referred to as R-Investors and their time- $t$  expectation of excess return from time  $t$  to time  $t+1$  is the true expectation. When S-Investors are imperfectly rational, they are referred to as I-Investors. They are also competitive in the same way as R-Investors. However, they are imperfectly rational in the sense that their expectation of  $Q_{t+1}$  at time  $t$  is biased due to irrational momentum sentiment

$$E_t^S [Q_{t+1}] = E_t [Q_{t+1}] + Z_t \quad (3.5)$$

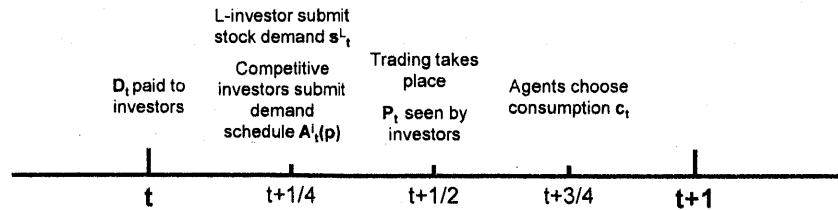
where  $Z_t$  is the bias at period  $t$ . Thus, as we shall see, their demand will be the sum of rational demand of R-Investors and the irrational component induced by the sentiment. In general, we let  $Z_t$  depend on the past realized excess returns:  $Z_t = Z(Q_{t-1}^e, Q_{t-2}^e, Q_{t-3}^e, \dots)$  where  $Q_{t-j}^e = Q_{t-j} - \bar{Q}$  with  $\bar{Q}$  being the long-run average of  $Q_t$ . For simplicity, we assume that  $Z(\cdot)$  is linear and depends on  $k$  lags of  $Q_t^e$ . Thus  $Z_t = b \underline{Q}_t$  where  $b$  is a  $1 \times k$  constant matrix and  $\underline{Q}_t = (Q_{t-1}^e \dots Q_{t-k}^e)^\top$ . In particular, if all entries in  $b$  are greater than 0, we have so-called "trend-chasing" or "momentum trading". The true expectation of  $Q_{t+1}$  is given by  $E_t [Q_{t+1}]$ , which is also what a fully rational investor expects. Suppose

that, in a particular realization,  $Q_{t-1} - \bar{Q} > 0$ . The excess return from  $t - 2$  to  $t - 1$  is higher than usual. The S-Investors observe this and they expect the future excess return for the next period ( $t$  to  $t + 1$ ) will be high as well. However, in our setup, we do not wish to let them be completely irrational. Thus, the expectation of the S-Investors still keeps the rational expectation part  $E_t[Q_{t+1}]$ . On top of that, we introduce a momentum or "trend-chasing" sentiment  $Z_t$ . Loosely speaking, they buy when price in the past few periods went up and sell when price went down. The strength and persistence of this momentum demand is given by the length and magnitude of  $b$ . This is similar to the positive feedback investors in Delong et al (1990), who have feedback demand on price gain in a 3-period model with no dividend. Since our model is capable of dealing with multiple lags of  $Q_t^e$ , it also covers a wide range of "technical rules" including momentum strategies. Other than the sentiment bias that I-Investors have about expectation of  $Q_{t+1}$ , they are completely rational in all other aspects. They have correct belief about how  $D_t$  evolves over time. Also, they have the right belief about the variance of excess return  $Var_t(Q_{t+1})$ . Note that R-Investors are a special subgroup of S-Investors with  $b = 0$ .

Previous literature on momentum or feedback trading have assumed that demand of momentum investors are exogeneously given by  $a + bQ_{t-1}^e$ <sup>4</sup>, which is inelastic of price. On the one hand, this greatly reduces the complexity. However, this may run into the danger of oversimplification. Dramatic results in these models may not hold when the momentum investors are still sensitive to price.

There is a single rational large investor (L-Investor). She is fully rational and her expectation is the same as the true expectation. In contrast with S-Investors, who individually consider themselves to be infinitesimally small, L-Investor is aware of the fact that she has price impact: if she buys, the execution price of the current period will become higher; if she sells, the execution price of the current period will become lower. Unlike competitive S-Investors, she does not submit demand schedule. Instead, she submits a market demand, which is inelastic of price. Given the demand schedule of the competitive investors, the L-Investor is effectively setting the price. She has the ability to move the price to a level that she desires.

### 3.2.3 Timeline of Trading



Events that take place during the period  $t$  to  $t + 1$  follow the sequence below:

---

<sup>4</sup>To be exact,  $a + b(P_{t-1} - P_{t-2})$ .

1. At time  $t + 0$ , dividend  $D_t$  is paid and every investor observes it.
2. At time  $t + \frac{1}{4}$ , the competitive investors submit demand schedule  $A_t^S(p)$  for the stock. The L-Investor submits an inelastic demand  $s_t^L$  for the stock.
3. At time  $t + \frac{1}{2}$ , trading takes place. Equilibrium price is determined so as to clear the market.
4. At time  $t + \frac{3}{4}$ , investors choose their consumption  $c_t^i$ ,  $i = S, L$ .

### 3.3 Equilibrium Definition and Solution

We shall solve the general S-Investors' dynamic optimization problem first, which gives us their demand schedule  $A_t^S(p)$ . Taking demand schedule as given, L-Investor can infer the price impact function and choose the market demand that fully optimizes dynamically. Given the infinite-horizon setup, we look for a stationary equilibrium, in which the strategies of investors are time-invariant.

#### 3.3.1 S-Investors' Problem

S-Investors are competitive. When they make trading decisions, they take the evolution of investment opportunities as given. In particular, the large investor's holding strategy and equilibrium price process are completely exogenous for them. Thus, they can formulate the process of excess return  $Q_{t+1}$  although their expectation is in general wrong<sup>5</sup>:

$$E_t^S [Q_{t+1}] = E_t [Q_{t+1}] + Z_t$$

Because each S-Investor is a price taker, she believes that she can buy or sell any amount of stock in the market at the prevailing price. Thus, we can define her total wealth  $W_t^S$  as the sum of her wealth in the riskless asset  $M_t^S$  and the "mark-to-market" value of her holding in the risky stock  $s_t^S P_t$ , where  $s_t^S$  is the number of shares she holds. Her wealth evolves as

$$W_{t+1}^S = (W_t^S - c_t^S) R + s_t^S Q_{t+1} \quad (3.6)$$

S-Investors solve the following optimization problem:

$$\begin{aligned} J_t^S &= \max_{c_t^S, s_t^S} E_t^S \left[ - \sum_{s=t}^{\infty} \beta^s e^{-\gamma^S c_s^S} \right] \\ \text{s.t. } W_{t+1}^S &= (W_t^S - c_t^S) R + s_t^S (Q_{t+1} + Z_t) \end{aligned} \quad (3.7)$$

---

<sup>5</sup>Only fully rational R-Investors' expectation is correct.

The solution to this maximization problem is similar to Wang (1994). We can solve for her optimal consumption  $c_t^S$ , optimal holding  $s_t^S$  of the risky stock under the equilibrium price process as well as her optimal demand schedule  $A_t^S(p)$  as a function of price. At time  $t + \frac{1}{4}$ , she submits  $A_t^S(p)$  and, under equilibrium price  $P_t$ , she gets  $s_t^S = A_t^S(P_t)$  number of shares.

### 3.3.2 L-Investor's Problem

L-Investor takes the demand schedule of competitive investors  $A_t^S(p)$  as given. She can solve for the clearing price  $P_t = F_t + \hat{P}_t(s_t^L)$  as a function of her inelastic market demand  $s_t^L$  from the market-clearing condition

$$A_t^S(p) \omega_S + s_t^L = \bar{\theta} \quad (3.8)$$

From the L-Investor point of view,  $F_t$  is completely exogenous but  $\hat{P}_t$  is a function of her demand  $s_t^L$ .  $\hat{P}_t(s_t^L)$  gives L-Investor the discount per share when demanding  $s_t^L$  units. Because of price impact, the L-Investor cannot mark the value of her stock holding to the market price. However, we can mark her stock holding to the fundamental value, which is exogeneous of her trading decision, and define her total wealth to be

$$W_t^L = M_t^L + s_{t-1}^L F_t \quad (3.9)$$

where  $M_t^L$  is her holding in riskless asset. We can interpret  $W_t^L$  as the sum of her wealth in riskless asset and the fundamental value of her stock holding at time  $t + \frac{1}{4}$  before trading takes place at  $t + \frac{1}{2}$ . It can be shown that

$$W_{t+1}^L = (W_t^L - c_t^L) R - (s_t^L - s_{t-1}^L) R \hat{P}_t(s_t^L) + s_t^L (1 + a) \varepsilon_{t+1} \quad (3.10)$$

Thus, she solves the following optimization problem

$$\begin{aligned} J_t^L &= \max_{c_t^L, s_t^L} E_t \left[ - \sum_{s=t}^{\infty} \beta^s e^{-\gamma^L c_s^L} \right] \\ \text{s.t. } W_{t+1}^L &= (W_t^L - c_t^L) R - (s_t^L - s_{t-1}^L) R \hat{P}_t(s_t^L) + s_t^L (1 + a) \varepsilon_{t+1} \end{aligned} \quad (3.11)$$

### 3.3.3 Solution of the Equilibrium

The equilibrium price process and the optimal policies can be expressed as functions of the state vector  $\Psi_t$ .  $\Psi_t$  has finite dimension and, together with  $W_t^i$  for  $i = S, L$ , are sufficient to characterize the state of the economy. With CARA utility, the investment policies do not depend on wealth of both S- and L-Investors. This makes price process independent of  $W_t^i$ . Thus, equilibrium price process and the optimal policies are function of  $\Psi_t$  only. Moreover, with proper choice of state variables and some augmentation,  $\Psi_t$  can be treated as Markov. We look for a stationary equilibrium given the horizon of the problem is infinite. So  $\hat{P}_t$ ,  $c_t^i, s_t^i$ ,  $i = S, L$  and  $A_t^S(p, D_t)$  are functions of  $\Psi_t$  only and do not depend on  $t$ . We can write  $P_t = F_t + \hat{P}(\Psi_t)$ ,  $c_t^i = c^i(\Psi_t)$ ,  $s_t^i = s^i(\Psi_t)$  and

$A_t^S = A^S(p, \Psi_t)$ . Given the CARA utility and Gaussian innovations we use, we strive to look for a linear equilibrium, in which  $\hat{P}(\cdot)$ ,  $c^i(\cdot)$ ,  $s^i(\cdot)$  and  $A^S(p, D_t, \cdot)$  (for a fixed  $p$ ) are all linear functionals of  $\Psi_t$ . Formally, we have

**Definition 1** *A stationary linear equilibrium is defined to be the stock price process  $P_t = F_t + \hat{P}(\cdot)$  and investor's' investment and consumption policies  $c^i(\cdot)$  ( $i = L, S$ ),  $s^L(\cdot)$ ,  $A^S(p, D_t, \cdot)$  that are linear in  $\Psi_t$  and independent of  $t$ , which satisfy*

1. The investors' policies maximize their utility
2. The stock market clears

We choose the state vector to be  $\Psi_t = (1 \quad s_{t-1}^L \quad L_t \quad Q_{t-1} \quad \dots \quad Q_{t-k})^\top$  where  $L_t = F_t + D_t - RP_{t-1}$  is the fundamental excess return. 1 is included for convenience of expression. The L-Investor's position before trading  $s_{t-1}^L$  is included because L-Investor cannot adjust her holding costlessly at the prevailing price. How much she already has before trading is a piece of important information.  $L_t$  is the fundamental excess return.  $Q_t = L_t + \hat{P}_t$  is determined by  $L_t$ , which is known at  $t$ , and  $\hat{P}_t$ , which is known at  $t + \frac{1}{2}$  after market-clearing price  $P_t$  is found.  $Q_t$  is an important piece of information as it will contribute to momentum sentiment bias going forward. Since  $Q_t$  is not determined when investors submit demand, we shall include the fundamental excess return  $L_t$  in  $\Psi_t$  instead.  $Q_{t-1}, \dots, Q_{t-k}$  are included since they determine the current momentum sentiment as well as future momentum sentiment.

Suppose that the L-Investor strategy is  $s_t^L = s_\Psi^L \Psi_t$  and price  $P_t = aD_t + P_\Psi \Psi_t$  for some  $1 \times (k+3)$  vectors  $s_\Psi^L$  and  $P_\Psi$ . Competitive investors can formulate the evolution of the state vector as a Gauss-Markov process

$$\Psi_{t+1} = a_\Psi \Psi_t + b_\Psi \varepsilon_{t+1} \quad (3.12)$$

where  $a_\Psi$  is a  $(k+3) \times (k+3)$  matrix and  $b_\Psi$  is a  $(k+3) \times 1$  vector. Subsequently, they can solve their dynamic programming problem.

**Proposition 1** *Given that  $s_t^L = s_\Psi^L \Psi_t$  and  $P_t = F_t + P_\Psi \Psi_t$ , the solution to the S-Investor optimization problem is as follows:*

$$\text{Value Function: } J_t^S(W_t^S, \Psi_t) = -\beta^t e^{-\alpha^S W_t^S - \frac{1}{2} \Psi_t' v^S \Psi_t} \quad (3.13)$$

where  $\alpha^S = \frac{\gamma^S r}{R}$  and  $v^S$  is a symmetric square matrix depending on  $s_\Psi^L, P_\Psi$ .

$$\text{Optimal Consumption : } c_t^S = -\frac{1}{\gamma^S} \ln \left( -\frac{1}{\gamma^S} \frac{\partial J_t^S}{\partial W_t^S} \right) \quad (3.14)$$

$$\text{Optimal Stock Holding : } s_t^S = s_\Psi^S \Psi_t \quad (3.15)$$

$$\text{Optimal Demand Schedule : } A_t^S(p, D_t, \Psi_t) = \phi \Psi_t + \rho(p - aD_t) \quad (3.16)$$

where  $s_\Psi^S$  and  $\phi$  are  $1 \times (k+3)$  vectors.  $\rho$  is a constant scalar.



At time  $t$ ,

$$\begin{aligned} -\ln(-J_{t+1}^S) &= \alpha^S W_{t+1}^S + \frac{1}{2} \Psi_{t+1}^\top v^S \Psi_{t+1} \\ &= \alpha^S W_{t+1}^S + \frac{1}{2} \Psi_t^\top a_\Psi^\top v^S a_\Psi \Psi_t + \Psi_t^\top a_\Psi^\top v^S b_\Psi \varepsilon_{t+1} + \frac{1}{2} b_\Psi^\top v^S b_\Psi \varepsilon_{t+1}^2 \end{aligned} \quad (3.17)$$

The first term is the log-utility due to wealth at  $t+1$ . The other three terms give us the log-utility due to future investment opportunities. In particular, the last term can be absorbed into density of  $\varepsilon_{t+1}$  by rescaling the variance. It can be shown that

$$E_t[J_{t+1}] \propto -E_t \left[ \beta^t e^{-\alpha^S W_{t+1}^S - \frac{1}{2} \Psi_t^\top a_\Psi^\top v^S a_\Psi \Psi_t - \Psi_t^\top a_\Psi^\top v^S b_\Psi \varepsilon_{t+1}} \right]$$

such that  $\varepsilon_{t+1} \sim N\left(0, (1/\sigma_D^2 + b_\Psi^\top v^S b_\Psi)^{-1}\right)$ . So  $\frac{1}{2} \Psi_t^\top a_\Psi^\top v^S a_\Psi \Psi_t$  gives the deterministic portion of investment opportunity and  $\Psi_t^\top a_\Psi^\top v^S b_\Psi \varepsilon_{t+1}$  gives the stochastic portion, which is linear in the shock  $\varepsilon_{t+1}$ . Define the investment opportunity as  $O_{t+1} = \frac{1}{2} \Psi_t^\top a_\Psi^\top v^S a_\Psi \Psi_t + \Psi_t^\top a_\Psi^\top v^S b_\Psi \varepsilon_{t+1}$ .

**Proposition 2** *The optimal stock holding of S-Investors can be rewritten as*

$$s_t^S = \frac{E_t(Q_{t+1} + Z_t) - \widetilde{Cov}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})} \quad (3.18)$$

where  $\widetilde{Cov}$  and  $\widetilde{Var}$  denotes the covariance and variance taken with rescaled distribution of  $\varepsilon_{t+1}$ .

Note that, when S-Investors are fully rational R-Investors,  $Z_t = 0$  and R-Investors' holding is

$$s_t^R = \frac{E_t(Q_{t+1}) - \widetilde{Cov}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})}$$

$\frac{E_t(Q_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})}$  is the myopic demand, which reflects the trade-off between expected return and risk.  $\frac{-\widetilde{Cov}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})}$  represents the hedging component. For a general S-Investor (possibly imperfectly rational), her demand is the sum of rational demand and irrational demand arising from momentum sentiment:

$$\begin{aligned} s_t^S &= \frac{E_t(Q_{t+1} + Z_t) - \widetilde{Cov}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})} \\ &= \underbrace{\frac{E_t(Q_{t+1}) - \widetilde{Cov}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{Var}_t(Q_{t+1})}}_{\text{Rational Demand}} + \underbrace{\frac{Z_t}{\alpha^S \widetilde{Var}_t(Q_{t+1})}}_{\text{Momentum Demand}} \end{aligned}$$

In the proof of Proposition 2, we also demonstrate the construction of competitive investors' optimal demand schedule as a function of price  $A^S(p, D_t, \Psi_t)$ . This allows us to calculate the L-Investor's price impact function from the market-clearing condition.

**Proposition 3**  $\hat{P}_t(s_t^L) = \pi_\Psi \Psi_t + \lambda s_t^L$  where  $\hat{\pi}_\Psi$  is a matrix of size  $1 \times (k+3)$  and  $\lambda$  is a scalar representing the price impact.

A well-defined price impact function allows L-Investor to fully optimize her trades dynamically. Her optimal strategy is given by the proposition below.

**Proposition 4** Given that price  $P_t = F_t + \hat{P}_t(s_t^L) = aD_t + \pi_\Psi \Psi_t + \lambda s_t^L$ . The solution to the L-Investor optimization is

$$\text{Value Function: } J_t^L(W_t^L, \Psi_t) = -\beta^t e^{-\alpha^L W_t^L - \frac{1}{2} \Psi_t^\top v^L \Psi_t} \quad (3.19)$$

where  $\alpha^L = \frac{\gamma^L r}{R}$  and  $v^L$  is a symmetric square matrix depending on  $\pi_\Psi$  and  $\lambda$ . Her optimal consumption is

$$c_t^L = -\frac{1}{\gamma^L} \ln \left( -\frac{1}{\gamma^L} \frac{\partial J_t^L}{\partial W_t^L} \right) \quad (3.20)$$

$$s_t^L = s_\Psi^L \Psi_t \quad (3.21)$$

where  $s_\Psi^L$  is a  $1 \times (k+3)$  vector depending on  $\pi_\Psi$  and  $\lambda$ .

### 3.4 Equilibrium Results

We are interested in studying how the non-competitive L-Investor trades against competitive S-Investors (with  $b \neq 0$ ). As discussed earlier, S-Investors can be perfectly rational R-Investors with  $b = 0$  or imperfectly rational I-Investors with  $b \neq 0$ . Also, to serve as a contrast, we can replace L-Investor with a group of competitive R-Investors. Thus, we have 4 cases to consider

1. Competitive R-Investors trading against competitive R-Investors
2. Competitive I-investors trading against competitive R-Investors
3. Competitive R-Investors trading against non-competitive L-Investor
4. Competitive I-Investors trading against non-competitive rational L-Investor

(1) is a special case of (2). (3) is a special case of (4). In the next 4 subsections, we shall study the above 4 cases one by one.

### 3.4.1 All Investors Are R-Investors

In this case, we have a group of R-Investors with total weight 2. Then, it is straightforward to show that

$$\begin{aligned} P_t &= -\frac{\alpha^R (1+a)^2 \sigma_D^2}{2r} + aD_t \\ s_t^R &= \frac{1}{2} \end{aligned}$$

Price consists of the fundamental value of dividend stream  $aD_t$  and the risk discount  $-\frac{\alpha^R (1+a)^2 \sigma_D^2}{2r}$ , which induces investors to hold  $\frac{1}{2}$  shares given that investors are risk-averse. Investors' holding is constant and, by symmetry, each agent should hold 1/2 since there is total weight of 2. Dividend shock is accommodated by change in equilibrium price since investors are homogeneous and there is no trading.

### 3.4.2 I-Investors Trading Against R-Investors

In this subsection, we shall study the case, in which general competitive S-Investors with potential bias ( $b \neq 0$ ) trade against perfectly rational R-Investors. Since both types are competitive, we do not need to find out their demand schedule explicitly. Their holding under equilibrium price will suffice. Thus, we can replace the fundamental excess return  $L_t$  by excess return  $Q_t$  in the state vector as, from the perspective of competitive investors, they view  $Q_t$  as determined rather than dependent on their holdings. In this case, the state vector is

$$\Psi_t = (1 \quad Q_t \quad \dots \quad Q_{t-k})^\top$$

The equilibrium excess return process follows an AR process

$$Q_{t+1} = \bar{\mu} + \mu_0 Q_t + \dots \mu_k Q_{t-k} + \varepsilon_{Q,t+1}. \quad (3.22)$$

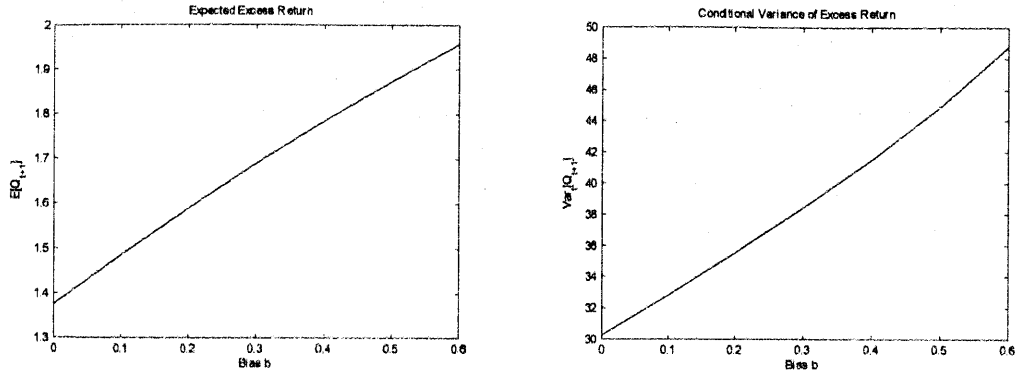
For simplicity, we shall consider one-dimensional bias  $b$ . This implies that, when forming expectation of  $Q_{t+1}$ , I-Investors' momentum sentiment is only based on the first lag of the deviation of realized excess return of last period from the unconditional mean,  $Q_{t-1} - \bar{Q}$ . With  $b > 0$ , we have momentum sentiment. In Fig 3.1, we plot expected return  $E[Q_t]$  and conditional variance  $Var_t[Q_{t+1}]$  as  $b$  increases.

With momentum depending on 1 lag only, the conditional variance of excess return is

$$Var_t[Q_{t+1}] = [P_{\Psi,L}(1+a) + (1+a)]^2 \sigma_D^2 \quad (3.23)$$

where  $P_{\Psi,L}$  is the coefficient in  $P_\Psi$  on state variable  $L_t$ . As  $b$  increases, the momentum sentiment is stronger. If the realized excess return of current period is unusually high, the momentum demand next period will be large. The rational portion of I-Investors and R-Investors' current demand would respond more aggressively to take advantage of the anticipated momentum bias next period. This pushes up the price for current period further. Thus,  $P_{\Psi,L}$  increases in  $b$ , which amplifies the fundamental risk and increases

Figure 3-1:  $E[Q_t]$  and  $Var_t[Q_{t+1}]$  for I-Investors v.s. R-Investors



Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^R = 1, \beta = 0.98$

the level of conditional variance of excess return as shown by the right panel of Fig 3.1. To bear higher conditional variance associated with stronger momentum bias, risk-averse investors require higher expected return to compensate the risk. This is illustrated by the left panel of Fig 3.1.

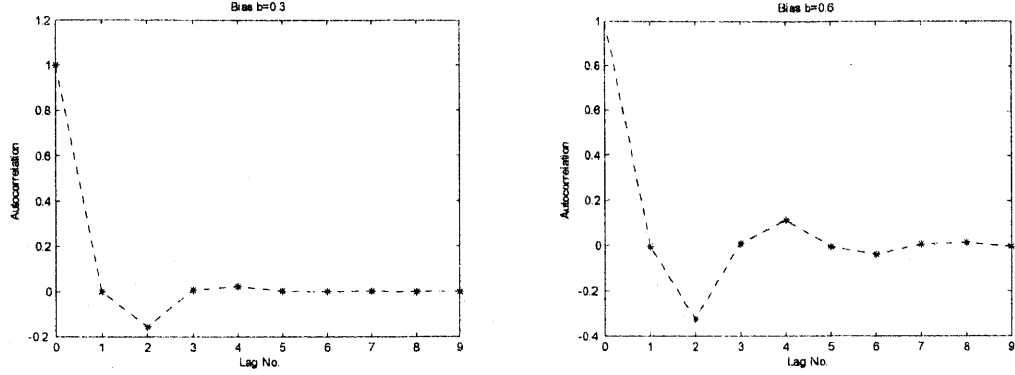
Fig 3.2 shows the autocorrelation of  $Q_t$  for 10 lags with the bias coefficient  $b = 0.3$  and  $0.6$ . Stronger momentum sentiment (more positive  $b$ ) leads to stronger reversal in actual excess return in equilibrium. This result is expected. Suppose that the realized excess return of last period was unusually high. This creates high momentum demand. To accomodate the momentum demand, rational demand must be less to clear the market. In order to induce rational demand to hold less of the risky security, the expected return for next period must be low. Thus, high past return leads to low future return<sup>6</sup>. Momentum trading cannot be self-fulfilling in equilibrium in the presence of fully rational investors.

### 3.4.3 R-Investors Trading Against L-Investor

In this subsection, we shall introduce the non-competitive L-Investor, who trades against fully rational R-Investors. If L-Investor were not to consider price impact, she would adjust her stock holding to the desired level immediately. With a non-competitive L-Investor, the situation is different. She will take her price impact into account and split her trading need into smaller orders to be executed over time. As a result, before submitting her market demand at time  $t$ , L-Investor will consider her previous period holding position  $s_{t-1}^L$ , which comes into the state vector as mentioned earlier:  $\Psi_t = (1 \ s_{t-1}^L \ L_t \ Q_{t-1} \ \dots \ Q_{t-k})^\top$ . Since all investors we consider here are rational ( $b = 0$ ) and they understand that past returns have no predictive power over future fundamental shocks, the coefficients on  $L_t$

<sup>6</sup>The adjacent excess returns are not correlated because momentum bias in expectation of  $Q_{t+1}$  depends on the excess return from time  $t - 2$  to  $t - 1$ .

Figure 3-2: Autocorrelation for I-Investors v.s. R-Investors



Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^R = 1, \beta = 0.98$

and  $Q_{t-1}, \dots, Q_{t-k}$ , which are introduced by momentum bias, are 0 in equilibrium for  $P_\Psi$ ,  $s_\Psi^R$ ,  $s_\Psi^L$  etc.

The L-Investor's trading strategy takes a simple form

$$s_t^L = s_{\Psi,1}^L + s_{\Psi,s}^L s_{t-1}^L \text{ with } 0 < s_{\Psi,s}^L < 1 \quad (3.24)$$

This implies that L-Investor trades gradually to reach the long-run equilibrium holding  $\bar{s}$ :  $(s_t^L - \bar{s}) = s_{\Psi,s}^L (s_{t-1}^L - \bar{s})$  where  $\bar{s} = \frac{s_{\Psi,1}^L}{1-s_{\Psi,s}^L}$ .  $\bar{s}$  can be interpreted as the long run equilibrium holding because  $\lim_{t \rightarrow \infty} (s_t^L - \bar{s}) = \lim_{t \rightarrow \infty} (s_{\Psi,s}^L)^t (s_0^L - \bar{s}) = 0$ . L-Investor's deviation from this long run limit at time  $t + \frac{1}{4}$  before trading is  $(s_{t-1}^L - \bar{s})$ . So her strategy is simply to hold  $s_t^L$  such that her deviation from  $\bar{s}$  will be a fraction  $s_{\Psi,s}^L$  of her previous deviation.

This suggests that L-Investor's orders are always in the same direction but the size decreases monotonically over time:

$$\begin{aligned} s_t^L - s_{t-1}^L &= (s_t^L - \bar{s}) - (s_{t-1}^L - \bar{s}) \\ &= (1 - s_{\Psi,s}^L) (s_{t-1}^L - \bar{s}) \\ &= (1 - s_{\Psi,s}^L) (s_{\Psi,s}^L)^{t-1} (s_0^L - \bar{s}) \end{aligned} \quad (3.25)$$

In the beginning, when her deviation from the long run equilibrium holding is large, she submits large orders to quickly reduce the deviation. The large price impact incurred is compensated by large trading need for optimal risk-return trade-off. This is similar to what is found by Almgren and Chriss (2000) and Vayanos (2001).

How fast she moves towards  $\bar{s}$  depends on  $s_{\Psi,s}^L$ . Smaller  $s_{\Psi,s}^L$  gives faster convergence. If L-Investor were to behave competitively and ignore price impact,  $s_{\Psi,s}^L = 0$ . She will move

to the long-run equilibrium holding instantly. But given that she takes her price impact into account,  $s_{\Psi,s}^L > 0$ . It takes many rounds of trading for her holding to reach this level asymptotically. It can be shown that:

**Proposition 5**  $s_{\Psi,s}^L$  is pinned down by the root  $\in [0, 1]$  of the following equation

$$0 = \frac{\alpha^R}{\alpha^L} (s_{\Psi,s}^L)^4 - \frac{\alpha^R}{\alpha^L} (s_{\Psi,s}^L)^3 - \left(1 + \frac{2\alpha^R}{\alpha^L}\right) R (s_{\Psi,s}^L)^2 + \left[\left(1 + \frac{2\alpha^R}{\alpha^L}\right) R + \frac{\alpha^R}{\alpha^L}\right] R s_{\Psi,s}^L - \frac{\alpha^R}{\alpha^L} R^2 \quad (3.26)$$

The speed of convergence to the long-run limit,  $\frac{1}{s_{\Psi,s}^L}$ , is decreasing in  $\frac{\alpha^R}{\alpha^L}$  and increasing in interest rate  $r$ . Furthermore, the long-run equilibrium holding by the large investor is given by

$$\bar{s} = \frac{\alpha^R}{\alpha^R + \alpha^L}$$

$\frac{\alpha^R}{\alpha^R + \alpha^L}$  is the optimal risk-sharing holding of the stock regardless of whether L-Investor is competitive or non-competitive. The more risk-averse L-Investor is, the faster she would move towards to  $\bar{s}$  as the cost of deviation from the risk-return optimal is larger. Since there is no hedging demand for competitive investors in this case, their demand will simply be

$$A_t^R(p) = \frac{E_t(P_{t+1} + D_{t+1}) - (1+r)p}{\alpha^R \sigma_Q^2} \quad (3.27)$$

The slope of the demand curve is  $\frac{dA_t^R}{dp} = -\frac{1+r}{\alpha^R \sigma_Q^2}$ . Higher effective risk aversion  $\alpha^R$  causes the coefficient on  $p$  to be smaller and gives rise to an inelastic demand. The immediate implication is that the price impact of L-Investor's trade is bigger. As a result, L-Investor tends to move slower to the long-run equilibrium position. Higher interest rate will increase the numerator directly and reduces the dividend risk in return  $\sigma_Q^2$  indirectly<sup>7</sup>. Both effects contribute to a more flatter demand curve, which leads to smaller price impact and faster movement towards optimal risk sharing. Thus, the speed of convergence is increasing in  $\alpha^L, r$  and decreasing in  $\alpha^R$ .

It is interesting to compare the price impact function generated from our equilibrium framework with exogenously specified ones in Almgren and Chriss (2000) and Stanzl and

<sup>7</sup> An upwards shock in dividend will increase current and expected future dividends. This gives a smaller shock to the fundamental value of stock  $F_t$  if the increases are discounted at a higher rate. Thus,  $\sigma_Q^2$  is decreasing in  $r$ .

Huberman (2005). In our framework, the price impact function is endogenously generated:

$$\begin{aligned}
P_t(s_t^L) &= aD_t + \pi_\Psi \Psi_t + \lambda s_t^L \\
&= aD_t + \pi_{\Psi,1} + \pi_{\Psi,s} s_{t-1}^L + \lambda s_t^L \\
&= aD_t + \pi_{\Psi,1} + (\pi_{\Psi,s} + \lambda) \left( s_0^L + \sum_{j=1}^{t-1} \Delta s_j^L \right) + \lambda \Delta s_t^L
\end{aligned} \tag{3.28}$$

where  $\Delta s_t^L = s_t^L - s_{t-1}^L$  is the order at time- $t$ . This resembles the exogeneously specified price impact function in the sense that the impact of trade is additive and linear. However, it is often assumed that there is a temporary price impact component due to transitory demand/supply imbalance, which goes away later on. As a result, the permanent price impact is, in general, smaller than the current period impact. In our equilibrium framework, there is no transitory demand/supply imbalance because investors are assumed to be always present. We find that the permanent price impact  $(\pi_{\Psi,s} + \lambda)$  is actually greater than current period impact  $\lambda$  since  $\pi_{\Psi,s}$  is positive. When L-Investor's pre-trading position  $s_{t-1}^L$  is high, R-Investors expect the L-Investor's holding for time- $t$  to be high as well since L-Investor trades gradually. Higher holding by L-Investor drives up expected price  $E_t[P_{t+1}]$ . As a result, their demand shifts upwards as represented by a positive  $\pi_{\Psi,s}$ .

Moreover, it is worth noting that, in equilibrium, the dividend volatility  $\sigma_D$  does not affect the trading speed. On the one hand, more volatile dividend implies more volatile price and excess return, which encourages L-Investor to trade more aggressively given that she is risk-averse. On the other hand, this makes the demand curve by R-Investors more inelastic, which increases the price impact. In equilibrium, these two effects cancel out exactly and  $\sigma_D$  does not affect the trading speed of L-Investor. This contrasts with result from Stanzl and Huberman (2005), which considers optimal trading in partial equilibrium and finds that price volatility increases aggressiveness of a large investor. The fact that demand curve becomes steeper and price impact becomes bigger is not captured in the partial equilibrium analysis.

### 3.4.4 I-Investors Trading Against L-Investor

In this subsection, we allow for general  $b \neq 0$ . We study how L-Investor trades against competitive S-Investors with irrational sentiment bias. We assume that with  $b = [b_1 \ b_2 \ \dots \ b_k]$ ,  $b_j > 0$  for  $j = 1, \dots, k$ . In section 4.3, we find that L-Investors takes price impact into consideration and trades gradually. Such gradual trading still carries over when momentum bias is present. But we will observe certain behaviors that we do not see when the other investors are fully rational.

**Proposition 6** *When a stationary equilibrium solution exists, there is always an eigenvector  $\bar{\Psi}$  of  $a_\Psi$  with corresponding eigenvalue being 1 such that  $\bar{\Psi} = E(\Psi_t)$ . Absent from further shocks, the system will converge to  $\bar{\Psi}$  asymptotically if started from some different state.*

This shows that, in general, there is always a steady state  $\bar{\Psi}$ , at which the system will stay unchanged if all future shocks realize to be 0. Furthermore, when the system deviates from the steady state, there is a natural tendency to for it to move back to the steady state absent of shocks.

L-Investor's trading strategy is  $s_t^L = s_{\Psi}^L \Psi_t$  where  $s_{\Psi}^L = (s_{\Psi,1}^L, s_{\Psi,s}^L, s_{\Psi,L}^L, s_{\Psi,Q_1}^L, \dots, s_{\Psi,Q_k}^L)^{\top}$ .  $s_{\Psi,s}^L$  is positive, which reflects the "gradual trading" effect carried over from the previous analysis. The interesting coefficients are  $s_{\Psi,L}^L$  and  $s_{\Psi,Q_1}^L, \dots, s_{\Psi,Q_k}^L$ , which are all 0 when L-Investor trades against fully rational R-Investors.  $s_{\Psi,L}^L$  is positive and becomes larger when  $b$  gets larger, which indicates greater effort by L-Investor to create momentum bias.  $s_{\Psi,Q_1}^L, \dots, s_{\Psi,Q_k}^L$  are negative as L-Investor wishes to take advantage of the mispricing created by the momentum bias.  $s_{\Psi,Q_1}^L, \dots, s_{\Psi,Q_{k-1}}^L$  become more negative as  $b$ 's magnitude increases initially. As  $b$  gets even larger, they turn around and become less negative. So their dependence on  $b$  is non-monotonic. This contrasts with the fact that the coefficient on the last lag of the realized excess return that is relevant for momentum demand,  $s_{\Psi,Q_k}^L$ , is always more negative when the magnitude of  $b$  increases. We will explain the reason behind it later. The L-Investor's trading strategy is fully dynamically optimizing. It will in general be hard to interpret why L-Investor trades in a certain way. Thus, we rely on impulse response analysis when a coordinate of  $\Psi_t$  is perturbed from the steady state. Her response to such an isolated shock will reveal her optimal trading strategy. We start with a scalar  $b$ , which allows the momentum bias to depend on the first lag of excess return only.

### Position Shock

Suppose that L-Investor's holding position is perturbed to be 0.1 share less than her steady state level at  $t = 1$ . For different magnitudes of  $b$ , Fig 3.3 shows L-Investor's optimal execution trajectories. When the bias coefficient  $b = 0$ , L-Investor's trading counterparties are fully rational. We observe gradual trading in consideration of price impact as analyzed in the section above. Her orders decline in size monotonically as deviation from desired holding level gets smaller. And her trades are all in the same direction. When  $b > 0$ , we see dramatic differences from this execution strategy. When  $b = 0.3$ , her order size no longer declines monotonically over time. She slows down trading after buying for 2 periods. This is because her buying in the first two periods pushed up momentum demand by I-Investors, which competes with her for buying the stock. L-Investor slows down her trading in consideration of momentum demand to avoid high execution cost and waits for momentum demand to decline. She then accelerates at  $t = 5$  to continue trading towards desired level. However, we need to note here, although the order size is no longer monotonically decreasing, the direction of orders<sup>8</sup> remains the same. When  $b$  gets even stronger ( $b = 0.5, 0.8$ ), we see round-trip trades along L-Investor's optimal holding trajectory. The direction of her trading is no longer the same, which contrasts with the cases when momentum bias  $b$  is small. She overbuys to pump up momentum demand and sells against it later on to take profit. This occurs for many cycles with declining amplitude until the momentum demand dies down. The price manipulation in

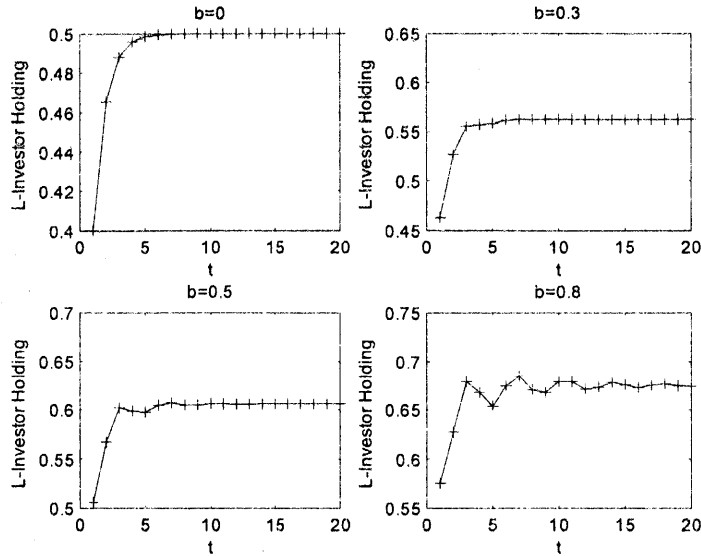
---

<sup>8</sup>i.e. buy or sell.



the latter cases ( $b = 0.5, 0.8$ ) is much more severe than in the case against rational R-Investors, where L-Investor only behaves non-competitively in a passive manner to avoid price impact and associated trading cost.

Figure 3-3: L-Investor's Holding Trajectories After A Position Shock

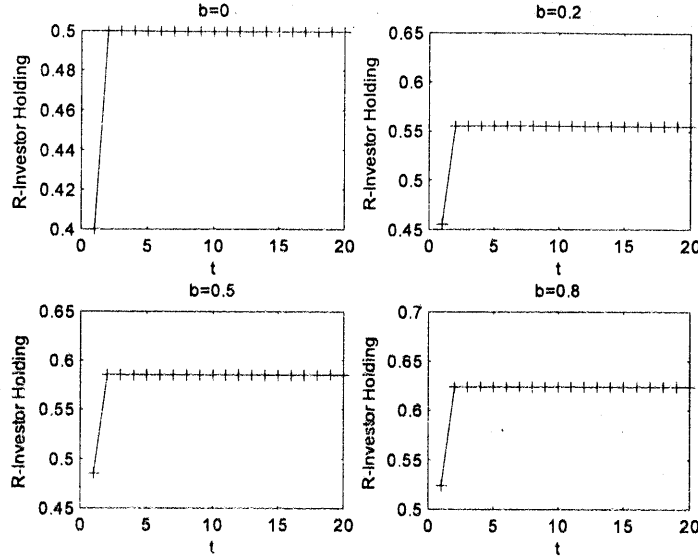


Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^L = 1, \beta = 0.98$

If we replace the L-Investor with R-Investors, they do not consider their holding for the last period when forming current period trading decision. They have no price impact and neither could they create any momentum through their trading. Thus, their holding returns to the steady state level immediately for all magnitudes of  $b$  as shown by Fig 3.4.

It is worth noting that oscillatory trading trajectory does not necessarily imply active price manipulation. When we let dividend shocks to be fully stochastic, we will observe oscillatory trading even when R-Investors trade with S-Investors. The distinction between this case and the L-Investor case is that R-Investors trade to take advantage of the oscillating momentum demand whereas L-Investor creates the momentum demand actively when bias coefficient is large. R-Investors' sell when price moves up due to positive momentum and buy when price goes down due to negative momentum. Thus, their trading tends to be in the opposite direction of price change. To create momentum, L-Investor exerts pressure on price and her trading tends to be in the same direction of price change. To illustrate the difference, we plot the correlation between trade and price changes in Fig 3.5 for both L-Investor and R-Investors trading against I-Investors. As we see from Fig 3.5, the correlation for L-Investor case is positive for all  $b$  whereas the correlation is negative for R-Investors. L-Investor contributes to the price changes on average. On the contrary, R-Investors' trade collectively to take advantage of price changes. With higher momentum coefficient  $b$ , R-Investors are more aggressive in taking advantage of

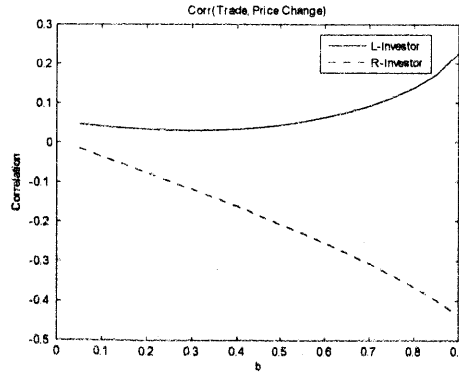
Figure 3-4: R-Investor's Holding Trajectories After A Position Shock



Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^R = 1, \beta = 0.98$

the mispricing and the correlation is more negative monotonically. In contrast, higher  $b$  leads L-Investor to take advantage of mispricing more aggressively. This generates lower correlation (albeit still positive) up to a level, beyond which correlation starts to increase drastically. This reflects the desire of L-Investor to intentionally manipulate the price when momentum strength is strong.

Figure 3-5:  $\text{Corr}(\text{Trade}_t, \Delta P_t)$



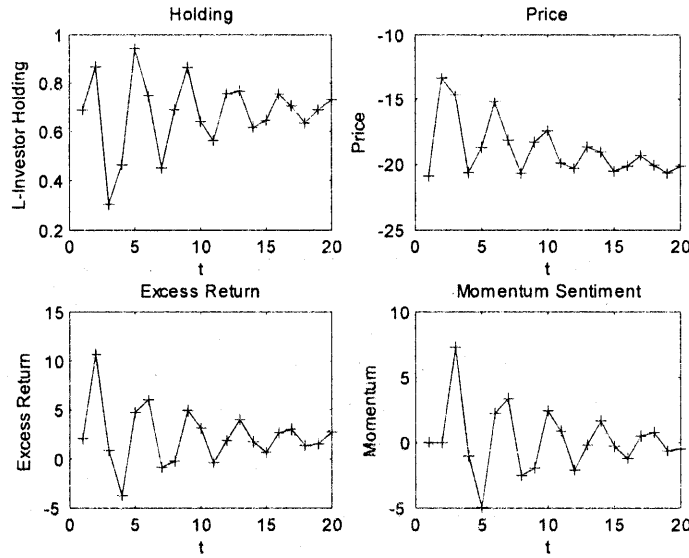
Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^L = \gamma^R = 1, \beta = 0.98$

### Dividend Shock

We shall next look at the effect of a dividend shock. Suppose, at  $t = 2$ , dividend shock  $\varepsilon_2 = 1$  for some realization and we let subsequent shocks be 0. For  $b = 0.85$  and  $a_D = 1$ ,

we have L-Investor's holding trajectory and equilibrium price path plotted in Fig 3.6. The shock takes place at time  $t = 2$ . At  $t = 2$ , the momentum bias is still 0. So I-Investors are still fully rational. At  $t = 3$ , the high realized excess return at  $t = 2$  generated by the shock will induce a positive momentum bias. However, price peaks at time  $t = 2$  rather than  $t = 3$ . Since, in this case, the dividend process is a random walk with no mean-reversion, there is no gradual return to 0 for the expected fundamental price. It might look puzzling why price becomes lower when the momentum demand is high at  $t = 3$ . This is due to the existence of rational demand. If the price were to rise further at  $t = 3$  due to the momentum demand, rational demand at  $t = 2$  would not miss this profitable opportunity and buy more at  $t = 2$ . In equilibrium, this would arbitrage away the predictable rise in price and price peaks at  $t = 2$ .

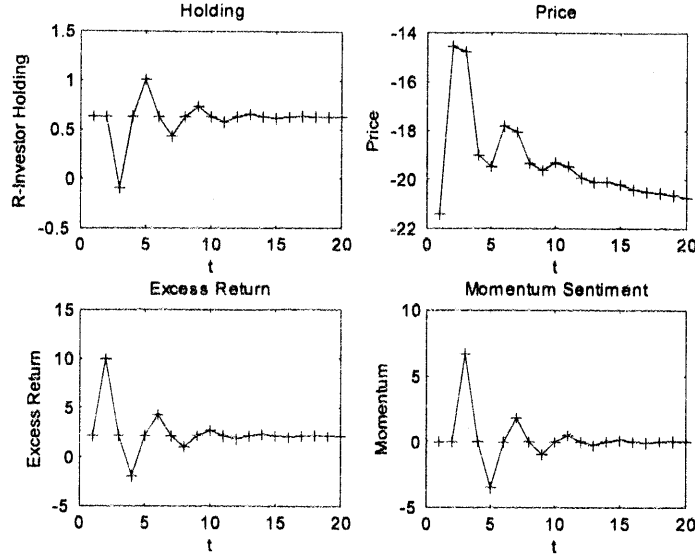
Figure 3-6: Trajectories After A Dividend Shock, With L-Investor



Parameters Used:  $b = 0.85, r = 0.1, a_D = 1, \sigma_D = 1, \gamma^I = \gamma^L = 1, \beta = 0.98$

What is interesting is the fact that L-Investor buys at  $t = 2$  and sells at  $t = 3$  at a slightly lower price while bearing the risk of holding on the stock and incurring price impact. Indeed, if it were R-Investors, who traded against I-Investors, their holding would not change at all at  $t = 2$  since all market participants are still rational at  $t = 2$ . To clear the market, the excess return from  $t = 2$  to  $t = 3$  must be  $\bar{Q}$  and price at  $t = 2$  must rise enough to offset the high demand to take advantage of the predictable momentum bias at  $t = 3$ . This is shown by Fig 3.7. L-Investor's purchase at time  $t = 2$  pushes the price up further to ensure a high realized excess return. This will induce higher momentum demand at  $t = 3$  and push the demand curve up at  $t = 3$ . L-Investor bears in mind that, at  $t = 3$ , she will wish to sell a lot to I-Investors to profit on their momentum demand. Thus, the shift in demand curve could benefit her tremendously with the large sale. The

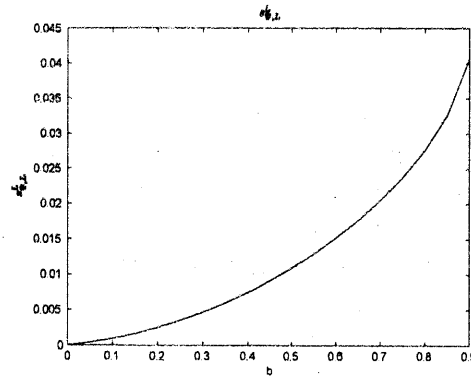
Figure 3-7: Trajectories After A Dividend Shock, With R-Investors



Parameters Used:  $b = 0.85, r = 0.1, a_D = 1, \sigma_D = 1, \gamma^I = \gamma^R = 1, \beta = 0.98$

benefit at  $t = 3$  outweighs the small loss from her initial purchase at  $t = 2$ . Effectively, she manipulates the intertemporal demand curve and benefits from such "reverse trading" scheme. To corroborate this interpretation, we plot how  $s_{\Psi, L}^L$ , the coefficient on  $L_t$  in L-Investor's trading strategy, changes with respect to  $b$  in Fig 3.8. If there is a dividend shock, the state variable it affects is  $L_t = (1 + a)D_t - RP_{t-1}$ . We can see that as the bias coefficient  $b$  increases, the coefficient on  $L_t$  rises at an increasing rate. Large coefficient on  $L_t$  will lead L-Investor to buy more after the dividend shock. Reverse trading will become more pronounced as  $b$  gets larger.

Figure 3-8: Coefficient on  $L_t$  in L-Investor's Strategy

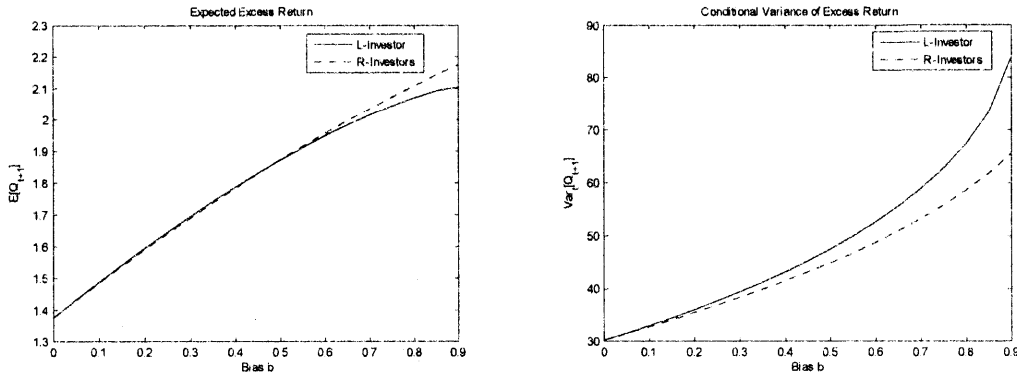


Parameters Used:  $r = 0.1, a_D = 1, \sigma_D = 1, \gamma^I = \gamma^L = 1, \beta = 0.98$

Besides, the system returns to steady state more quickly with R-Investors as they take advantage of the momentum sentiment aggressively and the momentum sentiment dies down fast. L-Investor is willing to take profit more moderately at each period. She aims to sustain momentum sentiment for a longer horizon. Price swings with much bigger magnitude and returns to the long-run stationary one at a slower speed, which presents more profitable opportunities. This can be interpreted as a "fishing in a common pool" problem. R-Investors are rational individuals. In equilibrium, they do not attempt to conserve the momentum sentiment and, therefore, "overfish". L-Investor, as a single large entity, tries to conserve the momentum sentiment and profits more.

### Expected Return and Volatility

Figure 3-9: Expected Excess Return and Conditional Variance



Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^L = \gamma^R = 1, \beta = 0.98$

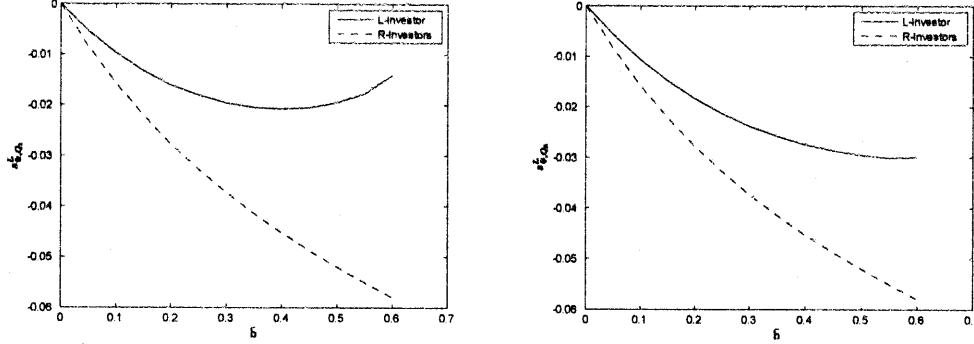
Fig 3.9 plots expected excess return and conditional variance of excess return for various levels of bias  $b$ . The conditional variance is significantly higher with L-Investor than with R-Investors. From (3.23), we know that conditional variance depends on  $P_{\Psi,L}$ , which measures the responsiveness of risk discount to fundamental excess return. As L-Investor exerts price pressure in the same direction of  $L_t$  to generate higher momentum bias,  $P_{\Psi,L}$  is larger with L-Investor, which contributes to a higher conditional variance. The difference in expected excess return is small relative to the difference in conditional variance.

### Multiple Lags

Finally, we shall examine momentum bias that depends on multiple lags of historical excess returns. The change of strategy coefficients on  $Q_{t-1}, \dots, Q_{t-k}$  as  $b$  gets larger is of particular interest. Let us fix  $k = 2$ . We let  $b = (\bar{b} \ \bar{b})$  and see how coefficients on  $Q_{t-1}$  and  $Q_{t-2}$  changes as  $\bar{b}$  increases. As shown by Fig 3.10, first of all, notice that coefficients are more negative for R-Investors than for L-Investor. This verifies our belief that R-Investors are more aggressive in taking advantage of the bias, which is positively related to past excess returns. Secondly, for R-Investors, both coefficients get more negative as  $\bar{b}$  increases. This reflects R-Investors' natural tendency to take advantage more aggressively when the bias becomes stronger. The coefficient on  $Q_{t-2}$ ,  $s_{\Psi,Q_2}^L$ , for L-Investor exhibits

the same pattern. As  $Q_{t-2}$  will not contribute to the momentum in the future, L-Investor takes more advantage of it when  $\bar{b}$  is bigger. However, the coefficient on  $Q_{t-1}$ ,  $s_{\psi, Q_1}^L$ , for L-Investor is not monotonic. While L-Investor wishes to take advantage the bias due to  $Q_{t-1}$ , she also wants to conserve the bias for future periods since, unlike  $Q_{t-2}$ ,  $Q_{t-1}$  will contribute to the momentum bias next period. As  $\bar{b}$  gets larger, the latter effect becomes stronger and the curve bends upwards.

Figure 3-10: Coefficients On Multiple Past Excess Returns



Parameters Used:  $r = 0.1, a_D = 0.9, \sigma_D = 1, \gamma^I = \gamma^L = \gamma^R = 1, \beta = 0.98$

### 3.5 Conclusion

In this paper, we study the optimal trading behavior of a risk-averse large institutional investor with price impact especially in the presence of momentum investors. In particular, we adopt a general equilibrium framework, which gives rise to endogenous price impact. In general, the large investor will split her trading need into smaller market orders over time. She trades more aggressively in the beginning when the deviation from her desired holding level is large and move gradually to desired holding position. In contrast with previous studies, we find price volatility does not increase her trading speed. This is because price impact becomes larger as her trading counterparties are also risk-averse. In the presence of momentum investors, the large investor engages in some interesting manipulation behaviors, which we do not observe without momentum demand. Depending on the strength of momentum trading, she may trade at non-monotonic speed or even engage in round-trip trades. Moreover, she conducts "reverse trading", by which she buys (sells) before a planned large sale (buy) to manipulate intertemporal demand curve. In addition, unlike competitive rational investors, who jump onto profitable mispricings produced by momentum aggressively, the large investor attempts to conserve the momentum sentiment so that it will last for longer period and present more profitable opportunities. Finally, our model suggests that momentum trading is not self-fulfilling. It leads to strong reversal in excess return.

## Appendix

### Proof of Proposition 2:

Define

$$\zeta_t = [Q_{t-1} \ Q_{t-2} \ \dots Q_{t-k}]^\top$$

The state vector is

$$\Psi_t = \begin{pmatrix} 1 \\ s_{t-1} \\ L_t \\ \zeta_t \end{pmatrix}$$

Notice that

$$\begin{aligned} L_{t+1} &= (1+a) D_{t+1} - R P_t \\ &= (1+a) D_{t+1} - R (a D_t + P_\Psi \Psi_t) \\ &= -R P_\Psi \Psi_t + (1+a) \varepsilon_{D,t+1} \end{aligned} \tag{3.29}$$

and

$$\begin{aligned} Q_t &= P_t + D_t - R P_{t-1} \\ &= (1+a) D_t + P_\Psi \Psi_t - R P_{t-1} \\ &= L_t + P_\Psi \Psi_t \end{aligned} \tag{3.30}$$

The evolution of  $\Psi_t$  follows

$$\Psi_{t+1} = \underbrace{\begin{pmatrix} e_1 \\ s_\Psi^L \\ -R P_\Psi \\ e_3 + P_\Psi \\ \vartheta \end{pmatrix}}_{a_\Psi} \Psi_t + \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1+a \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{b_\Psi} \varepsilon_{t+1} \tag{3.31}$$

where

$$\begin{aligned} e_j &= [0 \ 0 \ \dots 1 \ \dots 0] \text{ (1 at } j\text{th entry)} \\ \vartheta &= \begin{pmatrix} e_4 \\ e_5 \\ \vdots \\ e_{k+2} \end{pmatrix} \end{aligned}$$

The Bellman's equation is

$$\begin{aligned} J_t^S(W_t^S, \Psi_t) &= \max_{c_t^S, s_t^S} -\beta^t e^{-\gamma^S c_t^S} + E_t [J_{t+1}^S(W_{t+1}^S, \Psi_{t+1})] \\ \text{s.t. } W_{t+1}^S &= (W_t^S - c_t^S) R + s_t^S Q_{t+1} \end{aligned} \quad (3.32)$$

Note that because of the bias

$$\begin{aligned} E_t^S[Q_{t+1}] &= E_t[Q_{t+1}] + Z_t \\ Z_t &= b \bar{Q}_t^e \\ &= b \left[ \begin{pmatrix} \vartheta \\ e_{k+3} \end{pmatrix} - \begin{pmatrix} e_1 \\ \vdots \\ e_1 \end{pmatrix} \bar{Q} \right] \Psi_t \\ &= Z_\Psi \Psi_t \end{aligned}$$

As a result,

$$\begin{aligned} E_t^S[Q_{t+1}] &= E_t[(aD_{t+1} + P_\Psi \Psi_{t+1}) + D_{t+1} - R(aD_t + P_\Psi \Psi_t)] + Z_\Psi \Psi_t \\ &= [P_\Psi(a_\Psi - RI) + Z_\Psi] \Psi_t \\ &= e_\Psi^S \Psi_t \end{aligned} \quad (3.33)$$

$$\text{Var}_t^S[Q_{t+1}] = \text{Var}_t[Q_{t+1}] = [P_\Psi b_\Psi + (1+a)]^2 \sigma^2 \quad (3.34)$$

Conjecture that the value function takes the following form:

$$J_t^S(W_t^S, \Psi_t) = -\beta^t e^{-\alpha^S W_t^S - \frac{1}{2} \Psi_t^\top v^S \Psi_t} \quad (3.35)$$

Define  $v_{aa}^S = a_\Psi^\top v^S a_\Psi$ ,  $v_{bb}^S = b_\Psi^\top v^S b_\Psi$ ,  $v_{ab}^S = a_\Psi^\top v^S b_\Psi$ ,  $\Omega^S = (\Sigma^{-1} + v_{bb}^S)^{-1}$ ,  $\Gamma^S = (b_Q^S \Omega^S b_Q^S)^{-1}$ ,  $g^S = e_\Psi^S - b_Q^S \Omega^S v_{ab}^{S\top}$ ,  $d^S = |(\Omega^S)^{-1} \Sigma|^{-1/2}$ . Then

$$\begin{aligned} &E_t[J_{t+1}^S(W_{t+1}^S, \Psi_{t+1})] \\ &= -d^S \beta^{t+1} \exp \left[ -\alpha^S R(W_t^S - c_t) - \alpha^S \Psi_t^\top g^S s_t^S + \frac{1}{2} (\alpha^S)^2 (\Gamma^S)^{-1} (s_t^S)^2 \right. \\ &\quad \left. - \frac{1}{2} \Psi_t^\top (v_{aa}^S - v_{ab}^S \Omega^S v_{ab}^{S\top}) \Psi_t \right] \end{aligned}$$

F.O.C gives

$$s_t^S = \frac{1}{\alpha^S} f_\Psi^S \Psi_t \text{ where } f_\Psi^S = \Gamma^S g^S \quad (3.36)$$

$$c_t^S = \bar{c}^S + \frac{\alpha^S R}{\gamma^S + \alpha^S R} W_t^S + \frac{1}{2(\gamma^S + \alpha^S R)} \Psi_t^\top m^S \Psi_t \quad (3.37)$$

$$\begin{aligned} \text{where } \bar{c}^S &= \frac{1}{\gamma^S + \alpha^S R} \ln \left( \frac{\gamma^S}{\alpha^S \beta R d^S} \right) \\ m^S &= v_{aa}^S - v_{ab}^S \Omega^S v_{ab}^{S\top} + g^{S\top} \Gamma^S g^S \end{aligned}$$



Equating the two sides of the Bellman's equation, we have

$$\alpha^S = \frac{r\gamma^S}{R}, \bar{c}^S = -\frac{1}{\gamma^S R} \ln(r\beta d^S) \quad (3.38)$$

$$v^S = \frac{1}{R} m^S + 2 [\gamma^S \bar{c}^S + \ln(r/R)] \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} \quad (3.39)$$

To get demand schedule by competitive investors, we simply need to replace the equilibrium price  $P_t$  by  $p$ , the undetermined price level before market clearing. Define  $\Delta = p - aD_t$

$$\Psi_{t+1} = \underbrace{\begin{pmatrix} e_1 \\ s_\Psi^L 0 \\ -\text{Re}_{k+4} \\ e_3 + e_{k+4} \\ \vartheta \end{pmatrix}}_{\hat{a}_\Psi} + \underbrace{\begin{pmatrix} \Psi_t \\ \Delta \end{pmatrix}}_{\hat{\Psi}_t} + \begin{pmatrix} 0 \\ 0 \\ 1+a \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_{t+1} \quad (3.40)$$

Note that in the above equation we have  $e_j$  one dimension higher to accomodate  $\Delta$ . Also,

$$E_t^S [Q_{t+1}] = (P_\Psi \hat{a}_\Psi - \text{Re}_{k+4} + Z_\Psi) \hat{\Psi}_t = \hat{e}_\Psi^S \hat{\Psi}_t \quad (3.41)$$

With  $v^S$  calculated above already, we can find out competitive investors' demand schedule following similar optimization process

$$A^S = \begin{pmatrix} \phi & \rho \end{pmatrix} \hat{\Psi}_t \quad (3.42)$$

### Proof of Proposition 3:

Note that

$$\widetilde{\text{Cov}}_t(Q_{t+1}, O_{t+1}) = b_Q^S \Omega^S v_{ab}^{S\top} \quad (3.43)$$

$$\widetilde{\text{Var}}_t(Q_{t+1}) = (\Gamma^S)^{-1} \quad (3.44)$$

Thus,

$$\begin{aligned} s_t^S &= \frac{1}{\alpha^S} f_\Psi^S \Psi_t \\ &= \frac{E_t(Q_{t+1} + Z_t) - \widetilde{\text{Cov}}_t(Q_{t+1}, O_{t+1})}{\alpha^S \widetilde{\text{Var}}_t(Q_{t+1})} \end{aligned} \quad (3.45)$$

**Proof of Proposition 4:**

From market clearing condition,

$$\omega^S A^S + s_t^L = 1 \quad (3.46)$$

we get the price impact function for L-Investor

$$\begin{aligned} \pi_t(s_t^L) &= aD_t + \left( \frac{e_1 - \omega^S \phi}{\omega^S \rho} \right) \Psi_t - \frac{1}{\omega^S \rho} s_t^L \\ &= aD_t + \pi_\Psi \Psi_t + \lambda s_t^L \end{aligned} \quad (3.47)$$

where

$$\begin{aligned} \pi_\Psi &= \frac{e_1 \bar{\theta} - \omega^S \phi}{\omega^S \rho} \\ \lambda &= -\frac{1}{\omega^S \phi} \end{aligned}$$

**Proof of Proposition 5:**

For L-Investor, the evolution of her wealth is less straightforward. Unlike competitive investors, who can mark their holding of risk assets to the market price, she has to distinguish "paper wealth" and "liquidation wealth" as argued in Cherian and Jarrow (1993). As defined in earlier section,

$$\begin{aligned} W_{t+1}^L &= M_{t+1} + s_t^L a D_{t+1} \\ &= [M_t - c_t^L - (s_t^L - s_{t-1}^L) \pi_t(s_t^L)] R + s_t^L D_{t+1} + s_t^L a D_{t+1} \\ &= (M_t + a D_t - c_t^L) R + (s_{t-1}^L - s_t^L) (\pi_\Psi \Psi_t + \lambda H_t) R \\ &\quad - s_t^L (a R D_t - (1+a) a_D D_t - (1+a) \varepsilon_{t+1}) \\ &= (W_t^L - c_t^L) R + s_t^L (1+a) \varepsilon_{t+1} + \\ &\quad R \left( \Psi_t^\top e_2 \pi_\Psi \Psi_t + \lambda e_2^\top \Psi_t s_t^L - \pi_\Psi \Psi_t s_t^L - \lambda (s_t^L)^2 \right) \\ &= (W_t^L - c_t^L) R + s_t^L Q_{t+1}^L + \frac{1}{2} \Psi_t^\top w_\Psi^L \Psi_t \end{aligned} \quad (3.48)$$

where

$$\begin{aligned} w_\Psi^L &= e_2^\top \pi_\Psi + \pi_\Psi^\top e_2 \\ Q_{t+1}^L &= -R(\pi_\Psi - \lambda \tau_\Psi) \Psi_t - R \lambda s_t^L + (1+a) \varepsilon_{t+1} \\ &= e_\Psi^L \Psi_t + c_Q^L s_t^L + b_Q^L \varepsilon_{t+1} \end{aligned}$$

The state vector process  $\Psi_t$  is endogenous of her control variable  $s_t^L$ :

$$\begin{aligned}\Psi_{t+1} &= \underbrace{\begin{pmatrix} e_1 \\ 0 \\ -R\pi_\Psi \\ \pi_\Psi + e_3 \\ \vartheta \end{pmatrix}}_{a_\Psi^L} \Psi_t + \underbrace{\begin{pmatrix} 0 \\ 1 \\ -R\lambda \\ \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{c_\Psi^L} s_t^L + \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1+a \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{b_\Psi^L} \varepsilon_{t+1} \\ &= a_\Psi^L \Psi_t + c_\Psi^L s_t^L + b_\Psi^L \varepsilon_{t+1}\end{aligned}\quad (3.49)$$

The Bellman's equation is

$$\begin{aligned}J_t^L(W_t^L, \Psi_t) &= \max_{c_t^L, H_t} -\beta^t e^{-\gamma^L c_t^L} + E_t[J_{t+1}^L(W_{t+1}^L, \Psi_{t+1})] \\ \text{s.t. } W_{t+1}^L &= (W_t^L - c_t^L)R + s_t^L Q_{t+1}^L + \frac{1}{2} \Psi_t^\top w_\Psi^L \Psi_t \\ \Psi_{t+1} &= a_\Psi^L \Psi_t + c_\Psi^L s_t^L + b_\Psi^L \varepsilon_{t+1}\end{aligned}\quad (3.50)$$

Conjecture that the value function is of the form

$$J_t^L(W_t^L, \Psi_t) = -\beta^t e^{-\alpha^L W_t^L - \frac{1}{2} \Psi_t^\top v_\Psi^L \Psi_t}$$

Define  $v_{aa}^L = a_\Psi^{L\top} v^L a_\Psi^L$ ,  $v_{bb}^L = b_\Psi^{L\top} v^L b_\Psi^L$ ,  $v_{cc}^L = c_\Psi^{L\top} v^L c_\Psi^L$ ,  $v_{ab}^L = a_\Psi^{L\top} v^L a_\Psi^L$ ,  $v_{ca}^L = c_\Psi^{L\top} v^L a_\Psi^L$ ,  $v_{cb}^L = c_\Psi^{L\top} v^L b_\Psi^L$ ,  $\Omega^L = (\Sigma^{-1} + v_{bb}^L)^{-1}$ ,  $d^L = |\Sigma^{-1} \Omega^L|^{-1/2}$ ,  $g^L = \alpha^L e_\Psi^L + v_{ca}^L - (\alpha^L b_Q^L + v_{cb}^L) \Omega^L v_{ab}^{L\top}$ ,  $\Gamma^L = [(\alpha^L b_Q^L + v_{cb}^L) \Omega^L (\alpha^L b_Q^L + v_{cb}^L)^\top - 2\alpha^L c_Q^L - v_{cc}^L]^{-1}$   
Then

$$\begin{aligned}&E_t(J_{t+1}^L(W_{t+1}^L, \Psi_{t+1})) \\ &= -d^L \beta^{t+1} \exp \left[ -\alpha^L R (W_t^L - c_t^L) - g^L \Psi_t s_t^L + \frac{1}{2} (\Gamma^L)^{-1} (s_t^L)^2 \right. \\ &\quad \left. - \frac{1}{2} \Psi_t^\top \left( v_{aa}^L - v_{ab}^L \Omega^L v_{ab}^{L\top} + \alpha^L w_\Psi^L \right) \Psi_t \right]\end{aligned}\quad (3.51)$$

Therefore,

$$s_t^L = \Gamma^L g \Psi_t \quad (3.52)$$

$$c_t^L = \bar{c}^L + \frac{\alpha^L R}{\gamma^L + \alpha^L R} W_t^L + \frac{1}{2(\gamma + \alpha^L R)} m^L \Psi_t \quad (3.53)$$

where

$$\begin{aligned}\bar{c}^L &= \frac{1}{\gamma^L + \alpha^L R} \ln \left( \frac{\gamma^L}{\alpha^L \beta R d^L} \right) \\ m^L &= v_{aa}^L - v_{ab}^L \Omega^L v_{ab}^{L\top} + \alpha w_\Psi + g^{L\top} \Gamma^L g^L\end{aligned}$$

Equating the two sides of the Bellman's equation, we have

$$\alpha^L = \frac{r\gamma^L}{R}, \bar{c}^L = -\frac{1}{\gamma^L R} \ln(r\beta d^L) \quad (3.54)$$

$$v^L = \frac{1}{R} m^L + 2 [\gamma^L \bar{c}^L + \ln(r/R)] \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ \vdots & \dots & \dots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} \quad (3.55)$$

**Proof of Proposition 6:**

L-Investor's holding follows  $s_t^L = s_{\Psi,1}^L + s_{\Psi,s}^L s_{t-1}^L$  and price follows:  $P_t = aD_t + P_{\Psi,0} + P_{\Psi,s} s_{t-1}^L$ . Hence, state vector is

$$\Psi_t = \begin{pmatrix} 1 \\ s_{t-1}^L \end{pmatrix}$$

and it follows

$$\Psi_{t+1} = \begin{pmatrix} 1 & 0 \\ s_{\Psi,1}^L & s_{\Psi,s}^L \end{pmatrix} \Psi_t \quad (3.56)$$

which is deterministic.

Excess return for rational R-Investors is

$$Q_{t+1} = P_{\Psi,s} (s_{\Psi,1}^L - rP_{\Psi,0} \quad (s_{\Psi,s}^L - R)) \Psi_t + (1+a) \varepsilon_{t+1} \quad (3.57)$$

Since state vector at time  $t+1$  evolves deterministically, there is no hedging demand. It is well-known, for CARA utility function, the optimal demand can be found through the simple mean variance optimization.

$$\frac{E_t(Q_{t+1})}{\alpha^R \sigma_Q^2} = \frac{P_{\Psi,s} s_{\Psi,1}^L - rP_{\Psi,0}}{\alpha^R \sigma_Q^2} + \frac{P_{\Psi,s} (s_{\Psi,s}^L - R)}{\alpha^R \sigma_Q^2} s_{t-1}^L \quad (3.58)$$

Their demand schedule as a function of  $p$  is

$$A_t^R(p) = \phi_1 + \phi_s s_{t-1}^L + \rho(p - aD_t) \quad (3.59)$$

where

$$\phi_1 = \frac{P_{\Psi,1} + P_{\Psi,s} s_{\Psi,1}^L}{\alpha^R \sigma_Q^2}, \phi_s = \frac{P_{\Psi,s} s_{\Psi,1}^L}{\alpha^R \sigma_Q^2}, \rho = -\frac{R}{\alpha^R \sigma_Q^2}$$

Market clearing condition gives the price impact function

$$P(s_t^L) = aD_t + \pi_{\Psi,1} + \pi_{\Psi,s} s_{t-1}^L + \lambda s_t^L \quad (3.60)$$

where

$$\pi_{\Psi,1} = \frac{\phi_1 - 1}{\rho}, \pi_{\Psi,s} = \frac{\phi_s}{\rho}, \lambda = \frac{1}{\rho}$$

Large investor "market-to-fundamental value" wealth

$$W_{t+1} = (W_t - c_t) R - (s_t - s_{t-1}) R \hat{P}(s_t) + s_t \varepsilon_{Q,t+1}$$

Substitute this into L-Investor's value function

$$J_t^L(W_t^L, \Psi_t) = -\beta^t e^{-\alpha^L W_t^L - \frac{1}{2} \Psi_t^T v^L \Psi_t}$$

and take first order condition, we have L-Investor's optimal holding

$$s_t^L = \frac{-\alpha^L R \pi_{\Psi,1} + v_{12}^L}{\sigma_Q^2 ((\alpha^L)^2 + 2\alpha^L \alpha^R) - v_{22}^L} - \frac{\alpha^L R \left( \pi_{\Psi,s} - \frac{\alpha^R \sigma_Q^2}{R} \right)}{\sigma_Q^2 ((\alpha^L)^2 + 2\alpha^L \alpha^R) - v_{22}^L} s_{t-1}^L \quad (3.61)$$

where  $v_{12}, v_{22}$  satisfy

$$\begin{aligned} -\frac{(-\alpha^L R \pi_{\Psi,1} + v_{12}^L) \left( \alpha^L R \left( \pi_{\Psi,s} - \frac{\alpha^R \sigma_Q^2}{R} \right) \right)}{\sigma_Q^2 ((\alpha^L)^2 + 2\alpha^L \alpha^R) - v_{22}^L} + \alpha^L R \pi_{\Psi,1} &= R v_{12}^L \\ \frac{\left( \alpha^L R \left( \pi_{\Psi,s} - \frac{\alpha^R \sigma_Q^2}{R} \right) \right)^2}{\sigma_Q^2 ((\alpha^L)^2 + 2\alpha^L \alpha^R) - v_{22}^L} + 2\alpha^L R \pi_{\Psi,s} &= R v_{22}^L \end{aligned}$$

Market clearing requires

$$\frac{E_t(Q_{t+1})}{\alpha^R \sigma_Q^2} + s_t^L = 1$$

which gives us

$$P_{\Psi,s} = \frac{\alpha^R \sigma_Q^2}{R - s_{\Psi,s}^L} s_{\Psi,s}^L \quad (3.62)$$

$$P_{\Psi,1} = \left[ \frac{R}{R - s_{\Psi,s}^L} s_{\Psi,1}^L - 1 \right] \frac{\alpha^R \sigma_Q^2}{r} \quad (3.63)$$

This yields

$$\pi_{\Psi,s} = \frac{(s_{\Psi,s}^L)^2}{R (R - s_{\Psi,s}^L)} \alpha^R \sigma_Q^2 \quad (3.64)$$

$$\pi_{\Psi,1} = \left( \frac{\frac{1}{r} + \frac{s_{\Psi,s}^L}{R}}{R - s_{\Psi,s}^L} s_{\Psi,1}^L - \frac{1}{r} \right) \alpha^R \sigma_Q^2 \quad (3.65)$$

These give rise to the following equation that pins down  $s_{\Psi,s}^L$ :

$$\begin{aligned} & -\alpha^L R \left( \pi_{\Psi,s} - \frac{\alpha^R \sigma_Q^2}{R} \right) \pi_{\Psi,s} + 2\alpha^L R \pi_{\Psi,s} \\ & = R \left( \sigma_Q^2 \left( (\alpha^L)^2 + 2\alpha^L \alpha^R \right) + \alpha^L R \left( \pi_{\Psi,s} - \frac{\alpha^R \sigma_Q^2}{R} \right) \frac{1}{s_{\Psi,s}^L} \right) \end{aligned} \quad (3.66)$$

After substitution of  $\pi_{\Psi,s}$  and collecting the terms:

$$\varrho (s_{\Psi,s}^L)^4 - \varrho (s_{\Psi,s}^L)^3 - (1 + 2\varrho) R (s_{\Psi,s}^L)^2 + [(1 + 2\varrho) R + \varrho] R s_{\Psi,s}^L - \varrho R^2 = 0 \quad (3.67)$$

where  $\varrho = \frac{\alpha^R}{\alpha^L}$ . Note that the equation above have 4 roots. However, we look for a  $k_s$  such that  $|k_s| \leq 1$ . Otherwise, the large investor holding will go to infinity. Observe that when  $s_{\Psi,s}^L = 0$ , LHS =  $-\varrho R^2 < 0$ . When  $s_{\Psi,s}^L = 1$ , LHS =  $\varrho R r > 0$ . So there always exists a root between 0 and 1. Moreover, notice that when  $-1 < s_{\Psi,s}^L < 0$ ,

$$\begin{aligned} LHS &= \varrho (s_{\Psi,s}^L)^3 (s_{\Psi,s}^L - 1) + \left[ -(1 + 2\varrho) R (s_{\Psi,s}^L)^2 \right] + [(1 + 2\varrho) R + \varrho] R s_{\Psi,s}^L + (-\varrho R^2) \\ &\leq -2\varrho s_{\Psi,s}^L + \left[ -(1 + 2\varrho) R (s_{\Psi,s}^L)^2 \right] + [(1 + 2\varrho) R + \varrho] R s_{\Psi,s}^L + (-\varrho R^2) \\ &< 0 \end{aligned} \quad (3.68)$$

So there is no real root between -1 and 0.

The following equation pins down  $s_{\Psi,1}^L$ :

$$r P_{\Psi,1} - s_{\Psi,1}^L \left( P_{\Psi,s} - \frac{\sigma_Q^2 \alpha^R}{R} \right) \left( \frac{R}{s_{\Psi,s}^L} - 1 \right) = 0 \quad (3.69)$$

This yields  $s_{\Psi,1}^L = \frac{\varrho}{1+\varrho} (1 - s_{\Psi,s}^L)$ .

We can also show that  $s_{\Psi,s}^L$  is increasing in  $\varrho$ : Rearranging the original equation gives

$$\varrho \left[ (s_{\Psi,s}^L)^4 - (s_{\Psi,s}^L)^3 - 2R (s_{\Psi,s}^L)^2 + (2R + 1) R s_{\Psi,s}^L - R^2 \right] + R^2 s_{\Psi,s}^L - R (s_{\Psi,s}^L)^2 = 0 \quad (3.70)$$

implies  $\left[ (s_{\Psi,s}^L)^4 - (s_{\Psi,s}^L)^3 - 2R (s_{\Psi,s}^L)^2 + (2R + 1) R s_{\Psi,s}^L - R^2 \right] < 0$  since  $\varrho > 0$  and  $R^2 s_{\Psi,s}^L - R (s_{\Psi,s}^L)^2 > 0$  for  $0 \leq k_s \leq 1$ . If we raise  $\varrho$ , LHS will be less than 0 evaluated at the original root. Since RHS  $> 0$  at 1, this suggests that root has shifted to the right.

We can also show that  $s_{\Psi,s}^L$  is decreasing in  $R$ : Rearranging the original equation gives

$$\varrho \left( (s_{\Psi,s}^L)^4 - (s_{\Psi,s}^L)^3 \right) + R (R - s_{\Psi,s}^L) [(1 + 2\varrho) s_{\Psi,s}^L - \varrho] = 0 \quad (3.71)$$

If we raise  $R$ , LHS will be greater than 0 evaluated at the original root. This is because

$(1 + 2\varrho) s_{\Psi,s}^L - \varrho > 0$  at original root (since  $\varrho \left( (s_{\Psi,s}^L)^4 - (s_{\Psi,s}^L)^3 \right) < 0$  and  $R(R - s_{\Psi,s}^L)$  and  $R(R - s_{\Psi,s}^L)$  is increasing in  $R$  at original root. So the new root has shifted to the left. The above analysis also shows that  $k_s > \frac{\varrho}{2\varrho+1}$ .

**Proof of Proposition 7:**

Recall that  $\Psi_{t+1} = a_\Psi \Psi_t + b_\Psi \varepsilon_{t+1}$ . Taking expectation,  $E[\Psi_{t+1}] = a_\Psi E[\Psi_t]$ . Hence,  $\bar{\Psi} = a_\Psi \bar{\Psi}$ . Therefore,  $\bar{\Psi}$  is an eigenvector with associated eigenvalue 1. Furthermore, if the system is stationary, the other eigenvalues should be less than 1. Absent from further shocks, the system will converge to  $\bar{\Psi}$ .

## References

- Almgren, R. and N. Chriss 1999, "Optimal execution of portfolio transactions," *Journal of Risk*, 3, 5-39
- Aggarwal, R.K. and G. Wu 2006, "Stock market manipulations", *Journal of Business*, 79, 1915-1953
- Allen, F. and D. Gale 1992, "Stock-price Manipulation", *Review of Financial Studies*, 5, 503-529
- Bagnoli, M. and B. Lipman 1996, "Stock price manipulation through take-over bids", *RAND Journal of Economics*, 27, 124-147
- Barberis, N. and A. Shleifer 2003, "Style Investing," *Journal of Financial Economics*, 68, 161-199
- Benabou, R. and G. Laroque, 1992, "Using privileged information to manipulate markets: insiders, gurus, and credibility", *Quarterly Journal of Economics*, 107, 921-958
- Bertsimas, D. and A. Lo 1998, "Optimal control of execution costs," *Journal of Financial Markets*, 1, 1-50
- De Long, J.B., A. Shleifer, L. Summers and R. Waldman 1990, "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance*, 45, 379-395
- Hong, H., T. Lim and J. Stein 2000, "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies," *Journal of Finance*, 55, 265-295
- Goldstein, I. and A. Guebel 2008, "Manipulation and the allocational role of prices," *Review of Economic Studies*, 75, 133-164
- Jarrow, R. 1992, "Market manipulation, bubbles, corners, and short squeezes," *Journal of Financial and Quantitative Analysis*, 27, 311-336
- Kyle, A. 1985, "Continuous auction and insider trader," *Econometrica*, 53, 1315-1335
- Kyle, A. and S. Viswanathan 2008, "How to define illegal manipulation," *American Economic Review: Papers and Proceedings*, 98, 274-279
- Mei, J.P., G.J Wu and C.S Zhou 2004, "Behavior based manipulation: Theory and prosecution evidence," *Working Paper*
- Van Bommel, J. 2003, "Rumors," *Journal of Finance*, 58, 1499-1519
- Vayanos, D. 2001, "Strategic Trading in a dynamic noisy market," *Journal of Finance*, 56, 131-171
- Vila, J. 1989, "Simple games of market manipulation," *Economics Letters*, 29, 21-26
- Wang, J. 1994, "A model of competitive stock trading volume," *Journal of Political Economy* 102, 127-167